

Schriftliche Hausarbeit

Existence of quasi-stationary distributions for continuous-time Markov chains

Florian Ingerl

Datum der Abgabe 08.03.2019

Betreuer: Prof. Dr. Peter Pickl

Fakultät für Mathematik

Ludwig-Maximilians-Universität-München

Contents

1. Introduction	3
2. Continuous-time Markov chains	4
2.1. General theory and definitions	4
2.2. Transition function and q -matrix	7
2.3. Markov jump processes	10
2.4. Classification of states and stationary measures	17
2.5. Examples of continuous-time Markov chains	20
3. Renewal theory	21
3.1. Definitions	21
3.2. Iterates of the limit distribution of the remaining life time	21
3.3. Some convergence results	22
4. The Main Theorem	26
4.1. Setting and Theorem	26
4.2. A stationary Markov jump process related to qsd.	33
4.3. Quasi-stationary distribution as fixpoint	36
4.4. Checking the assumptions of Schauder-Tychonov fixpoint theorem	42
4.5. Associated renewal processes	45
4.6. Examples	50
4.6.1. The Random Walk	54
4.6.2. Linear birth and death process	54
4.6.3. M/M/ ∞ Queue	55
4.6.4. Markov branching process	56
5. Minimal quasi-stationary distributions	57
5.1. Setting and Main Theorem	57
5.2. Birth and Death Process	64
6. Acknowledgments	69
A. Appendix	69

1. Introduction

In this paper, we consider continuous time Markov chains $X = (X(t))_{t \geq 0}$ on the state space \mathbb{N}_0 with the origin 0 as an absorbing state and \mathbb{N} an irreducible class. It is further assumed that absorption is certain. We address the problem of the existence of so called quasi-stationary distributions, i.e. the set of initial distributions, which are such that the distribution of $X(t)$, conditioned upon non-absorption up to time t , is independent of t . These quasi-stationary distributions are interesting with respect to quasi-limiting distributions, i.e. the limit distribution of $X(t)$ as $t \rightarrow \infty$, conditioned upon non-absorption, which approximate the long time behaviour of a Markov process before it evanesces. In fact, we show that every quasi-limiting distribution must also be a quasi-stationary distribution.

A basic result is that the time of absorption or killing is exponential when starting from a quasi-stationary distribution, which implies that the rate of survival of a process must be at most exponential in order that a quasi-stationary distribution can exist, i.e. $E_x e^{\lambda R} < \infty$ for some $\lambda, x > 0$ where R is the absorption or killing time. The main target of this paper, is to show that this is also a sufficient condition for the existence of a quasi-stationary distribution under the additional assumption that the absorption time R with start in x goes to infinity in probability as $x \rightarrow \infty$. This result has been proven before in the special case of a birth and death process by van Doorn [19], by Ferrari, Martinez and Picco [11] and by van Doorn and Schrijner [20]. In our proof, we strongly orientate on the paper of Ferrari, Kesten, Martinez and Picco [10] and after its proof, we apply the theorem in a few examples.

As for uniqueness, it was pointed out by van Doorn [19] that for birth-and-death processes the set of QSD's under exponential killing can be a continuum or a singleton, see the example 4.6.1.

After that, we prove under additional constraints the existence of so called minimal quasi-stationary distributions, i.e. the quasi-stationary distributions taken as initial distributions where the mean time to absorption is minimal. We also prove that for a birth and death process, a unique minimal quasi-stationary distribution exists.

These results can be applied in the modelling of population sizes and in epidemics. A great monograph of quasi-stationary distributions and its applications is [8].

In the following, we give a broad preview to the individual chapters:

Chapter 2 serves as an introduction to Markov chains in continuous time both from a probabilistic view (i.e. we treat the construction of a minimal Markov jump process, strong Markov property, classification of states and stationary measures) and from an analytical view (i.e. we treat transition functions, Q-matrices and the Kolmogorov backward and forward equations) where our main references are Asmussen [3] and Anderson [2] respectively.

2. Continuous-time Markov chains

In Chapter 3, we review some basic facts about Renewal theory and present the results of Harkness and Shantaram [12] and a few technical Lemmas, that are needed to execute the proof of our main Theorem in Chapter 4.

Chapter 4 is the largest chapter and deals entirely with the proof of the Main Theorem (Theorem 4.6) and examples for its uses. A sketch for the proof is presented at the beginning of this chapter, that's why it is omitted here.

At last, Chapter 5 targets minimal quasi-stationary distributions under the additional restriction that the holding time parameters of the Markov chain are uniformly bounded. We also prove under these assumptions that for a birth-and-death process a unique minimal quasi-stationary distribution exists.

2. Continuous-time Markov chains

2.1. General theory and definitions

Although in the present paper, we are only concerned with Markov process with values in a countable state space (which we call then Markov chains), we will now derive the general theory for Markov processes with values in a polish space with the Borel- σ -algebra on it. So in this chapter let $(E, \mathcal{B}(E))$ be a polish space.

Definition 2.1. *A stochastic process $X = (X(t))_{t \in \mathbb{R}_{\geq 0}}$ defined on a measurable space (Ω, \mathcal{F}) with values in $(E, \mathcal{B}(E))$, is called a (continuous-time) Markov process with starting distributions $(P_x)_{x \in E}$ on (Ω, \mathcal{F}) if*

- (i) *For all $x \in E$ we have $P_x(X(0) = x) = 1$.*
- (ii) *The map $\kappa : E \times \mathcal{B}^{\otimes \mathbb{R}_{\geq 0}} \rightarrow [0, 1], (x, B) \mapsto P_x(X \in B)$ is a Markov kernel.*
- (iii) *The **weak Markov property** holds, i.e. for all $x \in E, s, t \geq 0$ and $A \in \mathcal{B}(E)$ holds*

$$P_x(X(t+s) \in A | \mathcal{F}_s) = P_{X_s}(X(t) \in A) := \kappa^t(X_s, A) \text{ a.s.} \quad (2.1)$$

*where $\mathcal{F}_t = \sigma\{X(s) : s \leq t\}$ is the canonical filtration associated with the process and where $\kappa^t : E \times \mathcal{B}(E) \rightarrow [0, 1], (x, A) \mapsto \kappa(x, \{y = (y(u))_{u \geq 0} : y(t) \in A\}) = P_x(X(t) \in A)$ is the **Markov kernel of the time difference t** .*

If E is a countable state space, then condition (ii) is always true, because every map from a space with the discrete σ -algebra on it is measurable.

By applying the tower property of the conditional expectation, one easily sees that the weak Markov property implies the **elementary Markov property**, i.e. for all $x \in E, s, t \geq 0$ and $A \in \mathcal{B}(E)$ holds

$$P_x(X(t+s) \in A | \mathcal{F}_s) = P_x(X(t+s) \in A | X(s)) \text{ a.s.} \quad (2.2)$$

2. Continuous-time Markov chains

In fact “weak Markov property=elementary Markov property+homogeneity in time”.

For the analysis of the sample paths of a Markov process, the weak Markov property of (2.1) is still inconvenient. For that reason, we have the following theorem.

Theorem 2.2. *The weak Markov property of (2.1) has an equivalent functional form, that is for all $\mathcal{B}(E)^{\otimes \mathbb{R}_{\geq 0}} - \mathcal{B}(\mathbb{R})$ measurable, bounded functions $f : E^{\mathbb{R}_{\geq 0}} \rightarrow \mathbb{R}$ and every $s \geq 0$ and $x \in E$ holds:*

$$E_x \left[f \left((X(t+s))_{t \in \mathbb{R}_{\geq 0}} \right) | \mathcal{F}_s \right] = E_{X(s)} [f(X)]. \quad (2.3)$$

Proof. “ \Leftarrow ” is trivial.

“ \Rightarrow ”: In the following argument, we make repeated use of the Monotone Class Theorem, see Theorem A.1. With the notation chosen as in the Monotone Class Theorem, we define the sets

$$\mathcal{H} = \{f : E^{\mathbb{R}_{\geq 0}} \rightarrow \mathbb{R} \text{ measurable} : \text{for all } s \geq 0, x \in E \text{ holds (2.3)}\}.$$

and

$$P = \{ \{y = (y(u))_{u \geq 0} : y(t_1) \in A_1, \dots, y(t_n) \in A_n\} | n \in \mathbb{N}, 0 \leq t_1 \leq t_2 \leq \dots \leq t_n < \infty, A_1, \dots, A_n \in \mathcal{B}(E) \}.$$

\mathcal{H} satisfies the conditions of the Monotone Class Theorem, because of linearity and dominated convergence of the conditional expectation. Because P is a \cap -stable generator of the σ -algebra $\mathcal{B}(E)^{\otimes \mathbb{R}_{\geq 0}}$, it suffices to prove for $A \in P$ that the indicator $\mathbf{1}_A$ is in \mathcal{H} . For this, we show the following claim by induction over n :

For all $n \in \mathbb{N}$ for all $0 \leq t_1 \leq \dots \leq t_n < \infty$, the set \mathcal{H} contains all $\mathcal{B}^{t_1, \dots, t_n}$ -measurable bounded functions (where $\mathcal{B}^{t_1, \dots, t_n}$ is the smallest σ -algebra on $E^{\mathbb{R}_{\geq 0}}$ such that the projections given from the exponent are measurable). For simplicity of notation, let $\Theta_s = (X(t+s))_{t \in \mathbb{R}_{\geq 0}}$.

n=1: So let $0 \leq t_1$ and $A_1 \in \mathcal{B}(E)$. Then owing to the weak Markov property, for all $s \geq 0, x \in E$ holds

$$\begin{aligned} E_x \left[\mathbf{1}_{\{y=(y(u))_{u \geq 0} : y(t_1) \in A_1\}} (\Theta_s) | \mathcal{F}_s \right] &= P_x (X(t_1+s) \in A_1 | \mathcal{F}_s) \\ &= P_{X(s)} (X(t_1) \in A_1) = E_{X(s)} \left[\mathbf{1}_{\{y=(y(u))_{u \geq 0} : y(t_1) \in A_1\}} (X) \right]. \end{aligned}$$

Then by the Monotone Class Theorem, \mathcal{H} contains all bounded $\mathcal{B}(E)^{t_1}$ -measurable functions.

n \mapsto n+1: So let $0 \leq t_1 \leq t_2 \leq \dots \leq t_{n+1}$ and $A_1, \dots, A_{n+1} \in \mathcal{B}(E)$. Then for all $s \geq 0, x \in E$ holds

$$\begin{aligned} &E_x \left[\mathbf{1}_{\{y=(y(u))_{u \geq 0} : y(t_1) \in A_1, \dots, y(t_{n+1}) \in A_{n+1}\}} (\Theta_s) | \mathcal{F}_s \right] \\ &= E_x \left[\mathbf{1}_{X(t_1+s) \in A_1} \cdots \mathbf{1}_{X(t_n+s) \in A_n} P_x (X(s+t_{n+1}) \in A_{n+1} | \mathcal{F}_{s+t_n}) | \mathcal{F}_s \right] \\ &= E_x \left[\mathbf{1}_{X(t_1+s) \in A_1} \cdots \mathbf{1}_{X(t_n+s) \in A_n} P_{X(s+t_n)} (X(t_{n+1}-t_n) \in A_{n+1}) | \mathcal{F}_s \right] \\ &= E_{X(s)} \left[\mathbf{1}_{X(t_1) \in A_1} \cdots \mathbf{1}_{X(t_n) \in A_n} P_{X(t_n)} (X(t_{n+1}-t_n) \in A_{n+1}) \right]. \end{aligned}$$

2. Continuous-time Markov chains

Now this last expression is a random variable that maps $\omega \in \Omega$ to:

$$\begin{aligned} \omega &\mapsto E_{X(s)(\omega)} \left[\mathbf{1}_{X(t_1) \in A_1} \cdots \mathbf{1}_{X(t_n) \in A_n} P_{X(s)(\omega)} (X(t_{n+1}) \in A_{n+1} | \mathcal{F}_{t_n}) \right] \\ &= E_{X(s)(\omega)} \left[\mathbf{1}_{\{y=(y(u))_{u \geq 0} : y(t_1) \in A_1, \dots, y(t_{n+1}) \in A_{n+1}\}} (X) \right]. \end{aligned}$$

Then by the Monotone Class Theorem, \mathcal{H} contains all bounded $\mathcal{B}(E)^{t_1, \dots, t_{n+1}}$ measurable functions, completing the induction step and proving the theorem. \square

Definition 2.3. We say that a Markov process $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ has the **strong Markov property** with respect to a stopping time τ , if we have a.s.

$$P_x (X(\tau + t) \in A | \mathcal{F}_\tau) \mathbf{1}_{\{\tau < \infty\}} = P_{X(\tau)} (X(t) \in A) \mathbf{1}_{\{\tau < \infty\}} \quad (2.4)$$

for all $x \in E, t \geq 0$ and $A \in \mathcal{B}(E)$.

We say that $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ is **strong Markov** if it has the strong Markov property w.r.t. any stopping time τ .

As with the weak Markov property, there is an equivalent functional form of the strong Markov property, i.e. (2.4) is equivalent to for all $\mathcal{B}(E)^{\otimes \mathbb{R}_{\geq 0}} - \mathcal{B}(\mathbb{R})$ measurable, bounded functions $f : E^{\mathbb{R}_{\geq 0}} \rightarrow \mathbb{R}$ and $x \in E$ holds:

$$E_x \left[f \left((X(\tau + t))_{t \in \mathbb{R}_{\geq 0}} \right) | \mathcal{F}_\tau \right] \mathbf{1}_{\{\tau < \infty\}} = E_{X(\tau)} [f(X)] \mathbf{1}_{\{\tau < \infty\}}. \quad (2.5)$$

The proof works similarly as above.

Theorem 2.4. A Markov process $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ has the strong Markov property with respect to any stopping time τ that charges only a countable number of values $(t_k)_{k \in \mathbb{N}} \cup \{\infty\}$.

Proof. At first, since τ does assume only a countable number of values, $X(\tau)$ is \mathcal{F}_τ measurable. So we need to show that for all $x \in E, t \geq 0, A \in \mathcal{B}(E)$ and $F \in \mathcal{F}_\tau$ we have

$$P_x (X(\tau + t) \in A, \tau < \infty, F) = E_x [P_{X(\tau)} (X(t) \in A), \tau < \infty, F].$$

If $F \subseteq \{\tau = t_k\}$, then this comes down to

$$\begin{aligned} P_x (X(\tau + t) \in A, \tau < \infty, F) &= E_x [P_x (X(t_k + t) \in A | \mathcal{F}_{t_k}) \mathbf{1}_F] \\ &= E_x [P_{X(t_k)} (X(t) \in A) \mathbf{1}_F] \\ &= E_x [P_{X(\tau)} (X(t) \in A), \tau < \infty, F]. \end{aligned}$$

In the general case, we decompose F into the disjoint union $F = \bigcup_{k \in \mathbb{N}} F \cap \{\tau = t_k\}$ and use σ -additivity. \square

The following theorem implies that any Markov jump process (defined later) is strong Markov.

Theorem 2.5. Let $X = (X(t))_{t \in \mathbb{R}_{\geq 0}}$ be a Markov process with right-continuous sample paths. Assume further that for any bounded and continuous function $f : E \rightarrow \mathbb{R}$ and any $s \geq 0$ the function $x \mapsto E_x [f(X_s)]$ is continuous. Then X is strong Markov.

2. Continuous-time Markov chains

Proof. So let τ be a stopping time. Since X has right-continuous paths $X(\tau)$ is \mathcal{F}_τ -measurable (see for instance Klenke [14] on page 455). We define for all $n \in \mathbb{N}$

$$\sigma_n = \sum_{k=1}^{\infty} \frac{k}{2^n} \mathbf{1}_{\{\frac{k-1}{2^n} < \tau \leq \frac{k}{2^n}\}}.$$

Then the σ_n 's are stopping times that charge only a countable number of values and obviously $\sigma_n \searrow \tau$.

Now let $x \in E, t \geq 0$ and $f : E \rightarrow \mathbb{R}$ bounded and continuous. If $F \in \mathcal{F}_\tau$, then $F \in \mathcal{F}_{\sigma_n}$ for all $n \in \mathbb{N}$ and because of Theorem 2.4, it holds

$$E_x[f(X(\sigma_n + t)), F] = E_x[E_{X(\sigma_n)}[f(X(t))], F].$$

Because X has right-continuous sample paths and f is continuous, we have $f(X(\sigma_n + t)) \xrightarrow{n \rightarrow \infty} f(X(\tau + t))$. Because of the assumption of this theorem, also $E_{X(\sigma_n)}[f(X(t))] \xrightarrow{n \rightarrow \infty} E_{X(\tau)}[f(X(t))]$ and thus dominated convergence gives

$$E_x[f(X(\tau + t)), F] = E_x[E_{X(\tau)}[f(X(t))], F].$$

We therefore found out that for all bounded and continuous $f : E \rightarrow \mathbb{R}$ holds

$$E_x[f(X(\tau + t)) | \mathcal{F}_\tau] = E_{X(\tau)}[f(X(t))].$$

Now for any closed set $A \subseteq E$ the indicator $\mathbf{1}_A$ can be approximated by bounded and continuous functions as follows: Let

$$\Phi : \mathbb{R} \rightarrow \mathbb{R}, t \rightarrow \begin{cases} 1, & \text{if } t \leq 0 \\ 1 - t, & \text{if } 0 \leq t \leq 1 \\ 0 & \text{else} \end{cases}.$$

and $f_\epsilon(x) = \Phi\left(\frac{\text{dist}(x, A)}{\epsilon}\right)$. We get with dominated convergence of the conditional expectation

$$\begin{aligned} P_x(X(\tau + t) \in A | \mathcal{F}_\tau) &= \lim_{\epsilon \rightarrow 0} E_x[f_\epsilon(X(\tau + t)) | \mathcal{F}_\tau] \\ &= \lim_{\epsilon \rightarrow 0} E_{X(\tau)}[f_\epsilon(X(t))] = P_{X(\tau)}(X(t) \in A). \end{aligned}$$

Since the closed sets generate the Borel- σ -algebra (recall that E was assumed Polish with the Borel- σ -algebra on it), the Monotone Class Theorem proves that X has the strong Markov property w.r.t. τ . \square

2.2. Transition function and q -matrix

From now on, we have the state space E countable (with the discrete topology on it).

2. Continuous-time Markov chains

To avoid misunderstandings due to notation, for any probability distribution μ on E we write P_μ for $P_\mu = \sum_{x \in E} \mu(x) P_x$ and in calculations where the starting distribution of the process does not matter, we just write P for the probability measure.

We define the **transition function** of a Markov process by

$$P_{ij}(t) = P_i(X(t) = j), i, j \in E, t \geq 0. \quad (2.6)$$

Because the distribution of a stochastic process, i.e. the probability of the events involving countably many of the random variables $X(t)$, is determined by its finite dimensional distributions, the distribution of a Markov process is determined by its starting distribution and its transition function. Indeed, for all $0 = t_0 \leq t_1 < t_2 < \dots < t_n$ and $i_1, \dots, i_n \in E$, we have

$$P(X(t_n) = i_n, \dots, X(t_1) = i_1) = \sum_{i_0 \in E} P(X(0) = i_0) \prod_{k=1}^n P_{i_{k-1}i_k}(t_k - t_{k-1})$$

as a simple application of Bayes' formula.

The transition function $P_{ij}(t)$ of a Markov process has the following properties:

- (1) $P_{ij}(t) \geq 0$ for all $i, j \in E, t \geq 0$ and $P_{ij}(0) = \delta_{ij}$.
- (2) $\sum_{j \in E} P_{ij}(t) = 1$ for all $t \geq 0, i \in E$.
- (3) For all $s, t \geq 0$ and $i, j \in E$, it follows from the Markov property $P_{ij}(s+t) = \sum_{k \in E} P_{ik}(s) P_{kj}(t)$, and these are called the Chapman-Kolmogorov equations.
- (4) If we are assuming our Markov process has right-continuous sample paths (with respect to the discrete topology on E of course), then we also have $\lim_{t \rightarrow 0} P_{ij}(t) = \delta_{ij}$.

Proof. For all $j \in E$, because the sample paths are right-continuous and $\{X(0, \omega)\}$ is a neighbourhood of $X(0, \omega)$ for all $\omega \in \Omega$, we have $\lim_{t \rightarrow 0} \mathbf{1}_{\{X(t)=j\}} = \mathbf{1}_{\{X(0)=j\}}$. Then since $0 \leq \mathbf{1}_{\{X(t)=j\}} \leq 1$, by the dominated convergence theorem, we get

$$\lim_{t \rightarrow 0} E[\mathbf{1}_{\{X(t)=j\}} | X(0) = i] = E[\mathbf{1}_{\{X(0)=j\}} | X(0) = i]$$

which is in other notation $\lim_{t \rightarrow 0} P_{ij}(t) = P_{ij}(0) = \delta_{ij}$. □

A function $P_{ij}(t)$ without reference to any Markov process, we call a transition function, if it satisfies properties (1),(3) and if $\sum_{j \in E} P_{ij}(t) \leq 1$ for all $t \geq 0, i \in E$. It is called **honest**, if it satisfies property (2) and **dishonest** otherwise. It is called **standard**, if it satisfies (4). In the following and in the literature we are referring to, the term transition function always means a possibly dishonest but standard transition function.

2. Continuous-time Markov chains

Anderson (see [2]) analyses transition functions in detail and in the following we shall establish his definitions and present some of the results of his book.

Proposition 2.6. *Let $P_{ij}(t)$ be a transition function. Then*

(1) $P_{ii}(t) > 0$ for all $t \geq 0$ and $i \in E$. If $i, j \in E$ with $i \neq j$, and if $P_{ij}(t) > 0$ for some $t > 0$, then $P_{ij}(t) > 0$ for all $t > 0$.

(2) If $P_{ii}(t) = 1$ for some $t > 0$, then $P_{ii}(t) = 1$ for all $t \geq 0$.

(3) If $t \geq 0$, then

$$|P_{ij}(t + \epsilon) - P_{ij}(t)| \leq 1 - P_{ii}(|\epsilon|)$$

and so $P_{ij}(t)$ is a uniformly continuous function of t .

Proof. See Anderson [2] Proposition 1.3. on page 7. □

Definition and Theorem 2.7. *Let $P_{ij}(t), i, j \in E, t \geq 0$ be a transition function on E . Then we call the matrix Q whose i, j -th component is the number $q(i, j) = P'_{ij}(0)$ the **q-matrix** of the transition function $P_{ij}(t)$. The diagonal components are non-positive and possibly infinite, the off-diagonal components are finite and non-negative, and the row sums are non-positive. A state $i \in E$ is called **stable** if $q(i, i)$ is finite and **instantaneous** otherwise. Q is called **stable** if all states $i \in E$ are stable and is further called **conservative**, if all row sums are equal to 0, that is if $\sum_{j \in E} q(i, j) = 0$ for all $i \in E$.*

Proof. See Anderson [2] pages 5-13. □

Proposition 2.8. *Let i be a stable state. Then*

(1) $P'_{ij}(t) \geq \sum_{k \in E} q(i, k) P_{kj}(t)$ for all $t \geq 0$ and $j \in E$.

(This is called the backward inequality.)

$P'_{ij}(t) \geq \sum_{k \in E} P_{ik}(t) q(k, j)$ for all $t \geq 0$ and $j \in E$.

(This is called the forward inequality.)

(2) If Q is conservative, then $P'_{ij}(t) = \sum_{k \in E} q(i, k) P_{kj}(t)$ for all $t \geq 0$ and $j \in E$.

Proof. See Anderson [2] Proposition 2.7 on page 13. □

The differential equations

$$P'_{ij}(t) = \sum_{k \in E} q(i, k) P_{kj}(t) \text{ for all } t \geq 0 \text{ and } j \in E,$$

which is in matrix notation just $P'(t) = QP(t)$, are called the **Kolmogorov backward equations** and $P'(t) = P(t)Q$ are called the **Kolmogorov forward equations**.

2. Continuous-time Markov chains

Definition 2.9. A square matrix $Q = (q(i, j))_{i, j \in E}$ is called a *q-matrix* if

$$\begin{aligned} 0 \leq q(i, j) < \infty \text{ for all } i, j \in E \text{ with } i \neq j, \\ \sum_{j \neq i} q(i, j) \leq q(i) \leq \infty \text{ for all } i \in E \text{ (where } q(i) = -q(i, i)). \end{aligned}$$

Q is called **stable** if $q(i) < \infty$, and **conservative** if equality holds in the second equation. A transition function $P_{ij}(t)$ is called a **Q-function** if $P'(0) = Q$.

Remark 2.10. The *q-matrix* has significant meaning for the theory of Markov processes. In fact, $q(i, j) = P'_{ij}(0)$ is equivalent to $P(X(t+h) = j | X(t) = i) = P_{ij}(h) = q(i, j)h + o(h)$. When attempting to model processes in real life by Markov processes, one usually has only $q(i, j)$ given from observations in a small time interval and may or may not be able to determine (maybe with the help of the Kolmogorov backward and forward equations) a possibly not unique transition function. That is the subject of the next theorem.

Theorem 2.11. Let Q be a stable *q-matrix*.

1. There exists a (possibly dishonest) transition function $f_{ij}(t)$ satisfying both the backward and forward Kolmogorov equations, with the property that $f_{ij}(t)$ is the minimal solution of each of these equations, in the sense that if $P_{ij}(t)$ is any other non-negative solution of either the forward or backward equation, then $f_{ij}(t) \leq P_{ij}(t)$ for all $i, j \in E, t \geq 0$. In particular $f_{ij}(t)$ is the minimal *Q-function*.
2. If $f_{ij}(t)$ is honest, then it is the unique solution to both the backward and forward equations, in the sense that there are no other non-negative solutions $P_{ij}(t)$ with $\sum_{j \in E} P_{ij}(t) \leq 1$, and in fact is the unique *Q-function*.

We call $f_{ij}(t)$ the **feller minimal Q-function** and a Markov process with transition function $f_{ij}(t)$ the **minimal process for Q** .

Proof. See Anderson [2] Theorem 2.2. on page 70. □

2.3. Markov jump processes

Definition 2.12. A Markov process $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ is called a **Markov jump process** if its sample paths are right-continuous and constant, except for isolated jumps.

Theorem 2.13. A Markov jump process $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ is strong Markov.

Proof. This is immediate from Theorem 2.5, because any function from the countable space E (which has the discrete topology) is continuous. □

Theorem 2.14. Let

$$J_n = \begin{cases} 0, & \text{if } n = 0 \\ \inf\{t > J_{n-1} | X(t) \neq X(t-)\}, & \text{if } n \geq 1 \end{cases}$$

denote the time of the n -th jump of $(X(t))_{t \in \mathbb{R}_{\geq 0}}$. Then the J_n are stopping times.

2. Continuous-time Markov chains

Proof. $J_0 = 0$ is a stopping time. Now assume, that for some $n \in \mathbb{N}_0$, J_n is a stopping time. For $t \geq 0$ define $Q(t) = \{qt : 0 \leq q \leq 1, q \in \mathbb{Q}\}$ and define

$$G = \bigcup_{u,s \in Q(t)} \{X(u) \neq X(s)\} \cap \{J_n \leq u, s\} \quad (2.7)$$

By induction hypotheses and because when taking countable unions, we remain in \mathcal{F}_t , we have $G \in \mathcal{F}_t$. Furthermore $G \subseteq \{J_{n+1} \leq t\}$ is clear. On the other hand, the paths in $\{J_{n+1} \leq t\}^c$ are constant on $[J_n, t]$, so

$$\{J_{n+1} \leq t\}^c \subseteq G^c = \bigcap_{u,s \in Q(t)} \{X(u) = X(s)\} \cup \{J_n > u\} \cup \{J_n > s\},$$

so $G = \{J_{n+1} \leq t\} \in \mathcal{F}_t$ and J_{n+1} is a stopping time. \square

Because J_n is a stopping time and X has right-continuous paths $X_n = X(J_n)$ is \mathcal{F}_{J_n} measurable (see for instance Klenke [14] on page 455) and in particular a random variable for all $n \in \mathbb{N}_0$. It is the state, just after the n -th jump. We call $(X_n)_{n \in \mathbb{N}_0}$ the **jump chain** or **embedded Markov chain**. Further we define the **sojourn or holding times** by $T_n = J_{n+1} - J_n, n \geq 0$.

Remark 2.15. *Two phenomena require our comment here. Firstly, a state $i \in E$ might be absorbing. So there is a last finite J_n and in this case we use the convention $T_n = T_{n+1} = \dots = \infty$ and $X_n = X_{n+1} = \dots = i$.*

*Secondly, jumps might congest, i.e. if $J_\infty = \sup_{n \in \mathbb{N}} J_n < \infty$. Then we say that J_∞ is the **time of first infinity** or **explosion time**. We have some thoughts about it later.*

Theorem 2.16. *Let $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, g \neq 0$ satisfy the Cauchy-exponential functional equation*

$$g(t+s) = g(t) \cdot g(s). \quad (2.8)$$

and assume further that g is either continuous or monotone. Then $g(t) = e^{Ct}$ for some arbitrary constant $C \in \mathbb{R}$. If further $g \leq 1$, then C must be smaller or equal 0.

Proof. See Theorem A.2 in the Appendix. \square

Theorem 2.17. *Let $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ be a Markov jump process. Then the joint distribution of the sequence $(X_n)_{n \in \mathbb{N}_0}$ of states visited (before explosion) and $(T_n)_{n \in \mathbb{N}_0}$ of holding times is given by*

1. $(X_n)_{n \in \mathbb{N}_0}$ is a Markov chain with transition matrix $P = (P(i, j))_{i, j \in E}$ given by $P(i, j) = P_i(X_1 = j)$.
2. There exists $q(i) \geq 0, i \in E$, such that, given $(X_n)_{n \in \mathbb{N}_0}$, the T_n are independent, with T_k being exponentially distributed with parameter $q(X_k)$.

2. Continuous-time Markov chains

Proof. Let \mathcal{B} denote the sets of the form $F_n = \{X_k = i_k, T_{k-1} > t_k, k = 1, \dots, n\}$ for some $i_1, \dots, i_n \in E$ and $t_1, \dots, t_n \geq 0$. Then \mathcal{B} together with the empty set, is a \cap -stable generator of $\sigma(X_n, n \geq 1, T_n, n \geq 0)$, so after the uniqueness theorem for measures, the joint distribution of Y_n, T_n is completely determined by the probabilities $P_i(B), i \in E, B \in \mathcal{B}$. Letting $i_0 = i$, the claim of the theorem is

$$P_i(F_n) = \prod_{k=1}^n P(i_{k-1}, i_k) e^{-q(i_{k-1})t_k} \quad (2.9)$$

for suitable $q(i)$.

We define

$$t + T_0 = \inf\{s > t : X(s) \neq X(s-)\}$$

and consider for $s, t \geq 0$

$$\begin{aligned} P_i(T_0 > t + s) &= E_i[P_i(T_0 > s, s + T_0 > t + s | \mathcal{F}_s)] \\ &= E_i[\mathbf{1}_{\{T_0 > s, X(s) = i\}} P_i(s + T_0 > t + s | \mathcal{F}_s)] \end{aligned}$$

Now let $B = \bigcap_{n \in \mathbb{N}} \{y \in E^{\mathbb{R}_{\geq 0}} : y(k \frac{t}{n}) = y(0) \text{ for } k = 0, 1, \dots, n\} \in (2^E)^{\otimes \mathbb{R}_{\geq 0}}$. Then because X has right-continuous sample paths and Theorem 2.2, we have

$$\begin{aligned} P_i(s + T_0 > t + s | \mathcal{F}_s) &= P_i(X(u) = X(s) \text{ for } u \in [s, s + t] | \mathcal{F}_s) \\ &= P_i((X(t + s))_{t \in \mathbb{R}_{\geq 0}} \in B | \mathcal{F}_s) \\ &= P_{X(s)}(X \in B) \\ &= P_{X(s)}(X(u) = X(0) \text{ for } u \in [0, t]) \\ &= P_{X(s)}(T_0 > t). \end{aligned}$$

So continuing the calculation, we have

$$\begin{aligned} P_i(T_0 > t + s) &= E_i[\mathbf{1}_{\{T_0 > s, X(s) = i\}} P_i(T_0 > t)] \\ &= P_i(T_0 > t) \cdot P_i(T_0 > s). \end{aligned}$$

Therefore the function $g(t) = P_i(T_0 > t)$ defined on $\mathbb{R}_{\geq 0}$ must satisfy the functional equation $g(t + s) = g(t) \cdot g(s)$. Since $g \neq 0, g \leq 1$ and g nonincreasing, $g(t) = e^{Ct}$ for some arbitrary constant $C \leq 0$. Thus T_0 must be exponentially distributed with parameter $q(i) = -C \geq 0$.

Next we show that the waiting time in a state and the next state are independent.

$$\begin{aligned} P_i(X_1 = j, T_0 > t) &= E_i[P_i(T_0 > t, X(t + T_0) = j | \mathcal{F}_t)] \\ &= E_i[\mathbf{1}_{\{T_0 > t, X(t) = i\}} \cdot P_i(X(t + T_0) = j | \mathcal{F}_t)] \end{aligned}$$

2. Continuous-time Markov chains

Now let $B = \bigcup_{q \in \mathbb{Q}_{>0}} \bigcap_{n \in \mathbb{N}} \{y \in E^{\mathbb{R}_{\geq 0}} : y(q) = j, y(k \frac{q}{n}) = y(0) \neq j \text{ for } k = 0, 1, \dots, n-1\} \in (2^E)^{\otimes \mathbb{R}_{\geq 0}}$. Then because X has right-continuous sample paths and Theorem 2.2, we have

$$\begin{aligned} P_i(X(t+T_0) = j | \mathcal{F}_t) &= P_i\left((X(s+t))_{s \in \mathbb{R}_{\geq 0}} \in B | \mathcal{F}_t\right) \\ &= P_{X(t)}(X \in B) \\ &= P_{X(t)}(X(T_0) = j). \end{aligned}$$

So continuing the calculation, we have

$$\begin{aligned} P_i(X_1 = j, T_0 > t) &= E_i[\mathbf{1}_{\{T_0 > t, X(t)=i\}} \cdot P_i(X(T_0) = j)] \\ &= P_i(X_1 = j) \cdot P_i(T_0 > t) \\ &= P(i, j) e^{-q(i)t}. \end{aligned}$$

This is (2.9) for $n = 1$.

The case $n > 1$ follows by the **strong** Markov property (or more precisely its functional equivalent) and induction over n . Define $J_{n-1} + T_0 = \inf\{s > J_{n-1} : X(s) \neq X(s-)\}$, then

$$\begin{aligned} P_i(F_n) &= E_i[P_i(F_{n-1}, X(J_{n-1} + T_0) = i_n, J_{n-1} + T_0 > t_n + J_{n-1} | \mathcal{F}_{J_{n-1}})] \\ &= E_i[\mathbf{1}_{\{F_{n-1}, X(J_{n-1})=i_{n-1}\}} \cdot P_i(X(J_{n-1} + T_0) = i_n, J_{n-1} + T_0 > t_n + J_{n-1} | X(J_{n-1}))] \\ &= E_i[\mathbf{1}_{\{F_{n-1}\}} \cdot P_{i_{n-1}}(X(T_0) = i_n, T_0 > t_n)] \\ &= P_i(F_{n-1}) \cdot P_{i_{n-1}}(X(T_0) = i_n, T_0 > t_n). \end{aligned}$$

□

Given a stochastic matrix $P = (P(i, j))_{i, j \in E}$ on E and holding times $q(i) \geq 0, i \in E$, can we construct a Markov jump process, whose embedded Markov chain has transition matrix given by P and whose holding times are given by the $q(i)$? It is indeed possible, and we will describe here the so called **minimal construction**. Let $\Delta \notin E$ be some extra state (to describe the process after explosions) and define $E_\Delta = E \cup \Delta$ and $q(\Delta) = 0, P(\Delta, \Delta) = 1$. Let the sample space be

$$\Omega = E_\Delta \times \overline{\mathbb{R}} \times E_\Delta \times \overline{\mathbb{R}} \dots$$

and let $X_0, T_0, X_1, T_1, X_2, \dots$ be the obvious coordinate functions. With the help of the Theorem of Ionescu-Tulcea (see for instance Klenke [14] Theorem 14.32. on page 286), it is possible to construct probabilities $P_i, i \in E$ on $(\Omega, 2^{E_\Delta} \otimes \mathcal{B}(\overline{\mathbb{R}}) \otimes 2^{E_\Delta} \otimes \dots)$ such that:

1. $(X_n)_{n \in \mathbb{N}_0}$ is a Markov chain with transition matrix P and $P_i(X_0 = i) = 1, i \in E$.
2. Given (X_n) , the T_n are independent with T_k being exponentially distributed with intensity $q(X_k)$.

2. Continuous-time Markov chains

Set $J_0 = 0, J_n = T_0 + \dots + T_{n-1}$, $\omega(\Delta) = \sup_{n \in \mathbb{N}} J_n$ and define

$$X(t) = \begin{cases} X_k & \text{if } J_k \leq t < J_{k+1} \\ \Delta & \text{if } t \geq \omega(\Delta) \end{cases}$$

Theorem 2.18. $(X(t))_{t \geq 0}$ is a Markov jump process on $E \cup \Delta$, whose embedded Markov chain has transition matrix given by P and whose holding time exponential parameters are given by the $q(i)$.

Proof. From the construction it is clear that $(X_t)_{t \geq 0}$ has right-continuous sample paths. Also if we show that $(X_t)_{t \geq 0}$ is a Markov process (and hence a Markov jump process), then it is clear from the construction that $(X_n)_{n \in \mathbb{N}_0}$ is the embedded Markov chain of $(X_t)_{t \geq 0}$ and that the holding time parameters are given by the $q(i)$.

Thus we only need to verify the Markov property, i.e. for all $s, t \geq 0, z, y \in E \cup \Delta$ holds

$$P_z(X_{t+s} = y | \mathcal{F}_t) = P_{X_t}(X_s = y). \quad (2.10)$$

Of course, we have

$$P_z(X_{t+s} = y | \mathcal{F}_t) \mathbf{1}_{\{t \geq \omega(\Delta)\}} = \delta_{y\Delta} \mathbf{1}_{\{t \geq \omega(\Delta)\}} = P_{X_t}(X_s = y) \mathbf{1}_{\{t \geq \omega(\Delta)\}}$$

so it suffices to verify (2.10) on $\{t < \omega(\Delta)\}$, so we operate on $\{t < \omega(\Delta)\}$ for the rest of the proof. Then we can define the waiting time until the next jump at time t for $t \geq 0$ by

$$R_t = J_{n(t)} - t, \text{ where } n(t) = \min\{n : J_n > t\}.$$

We also define the random vector

$$M_t = (T_{n(t)}, T_{n(t)+1}, \dots, X_{n(t)-1}, X_{n(t)}, \dots).$$

Then $(X_{t+s})_{s \geq 0}$ is constructed from (R_t, M_t) in the same way as $(X_t)_{t \geq 0}$ is constructed from

$$(R_0, M_0) = (T_0, M_0) = (T_0, T_1, \dots, X_0, X_1, \dots).$$

Indeed, we have by construction

$$X_t = \begin{cases} X_0 & \text{if } 0 \leq t < T_0 \\ X_k & \text{if } T_0 + \dots + T_{k-1} \leq t < T_0 + \dots + T_k, k \geq 1 \\ \Delta & \text{if } t \geq \sum_{n \in \mathbb{N}_0} T_n \end{cases}$$

and on the other hand

$$X_{t+s} = \begin{cases} X_{n(t)-1} & \text{if } 0 \leq s < J_{n(t)} - t \\ X_{n(t)} & \text{if } J_{n(t)} - t \leq s < J_{n(t)} - t + T_{n(t)} \\ X_{n(t)-1+k} & \text{if } J_{n(t)} - t + \dots + T_{n(t)+k-2} \leq s < J_{n(t)} - t + \dots + T_{n(t)+k-1}, k \geq 2 \\ \Delta & \text{if } s \geq J_{n(t)} - t + \sum_{k \in \mathbb{N}_0} T_{n(t)+k} \end{cases}$$

2. Continuous-time Markov chains

Hence, it suffices to show the conditional distribution of (R_t, M_t) given \mathcal{F}_t is the P_{X_t} -distribution of (R_0, M_0) . Now $((X_n, T_n))_{n \in \mathbb{N}_0}$ is a Markov chain with state space $E \cup \Delta \times (0, \infty]$ and transition kernel given by

$$P(X_{n+1} = j, T_{n+1} > t | \mathcal{H}_n) = P(X_n, j) e^{-q(j)t}$$

with the natural filtration $\mathcal{H}_n = \sigma(X_k, T_k : k \leq n)$. Since for all $n \in \mathbb{N}_0$ we have $\{n(t) - 1 = n\} = \{T_0 + \dots + T_{n-1} \leq t < T_0 + \dots + T_n\} \in \mathcal{H}_n$ it follows that $n(t) - 1$ is a stopping time w.r.t this chain. Also $\mathcal{F}_t \subseteq \mathcal{H}_{n(t)-1}$ because \mathcal{F}_t is generated by all sets of the form $\{X_{s_1} = y_1, \dots, X_{s_m} = y_m\}, m \in \mathbb{N}, 0 \leq s_1 < \dots < s_m \leq t, y_1, \dots, y_m \in E \cup \Delta$ and for all $n \in \mathbb{N}_0$ we have

$$\begin{aligned} & \{X_{s_1} = y_1, \dots, X_{s_m} = y_m\} \cap \{n(t) - 1 = n\} \\ &= \bigcup_{k_1=0}^n \bigcup_{k_2=k_1}^n \dots \bigcup_{k_m=k_{m-1}}^n \{J_{k_1} \leq s_1 < J_{k_1+1}, \dots, J_{k_m} \leq s_m < J_{k_m+1}, X_{k_1} = y_1, \dots, X_{k_m} = y_m\} \\ & \cap \{J_n \leq t < J_{n+1}\} \in \mathcal{H}_n. \end{aligned}$$

Furthermore R_t is $\mathcal{H}_{n(t)-1}$ -measurable, because for all $s \geq 0$ and $n \in \mathbb{N}_0$ we have

$$\{R_t > s\} \cap \{n(t) - 1 = n\} = \{J_{n+1} > t + s\} \cap \{J_n \leq t < J_{n+1}\} \in \mathcal{H}_n$$

and $X_{n(t)-1} = X_t$, because

$$X_{n(t)-1} = \sum_{k \in \mathbb{N}_0} X_k \mathbf{1}_{\{n(t)-1=k\}} = \sum_{k \in \mathbb{N}_0} X_k \mathbf{1}_{\{J_k \leq t < J_{k+1}\}} = X_t.$$

Thus we have for all $m \in \mathbb{N}$ and $s, s_1, \dots, s_m \geq 0, y_0, y_1, \dots, y_m \in E \cup \Delta$ with the tower-property of the conditional expectation and the strong Markov property

$$\begin{aligned} & P_z(R_t > s, T_{n(t)} > s_1, T_{n(t)+1} > s_2, \dots, T_{n(t)+m-1} > s_m, X_{n(t)-1} = y_0, \dots, X_{n(t)-1+m} = y_m | \mathcal{F}_t) \\ &= E_z[\mathbf{1}_{\{R_t > s\}} P_z(T_{n(t)} > s_1, \dots, T_{n(t)+m-1} > s_m, X_{n(t)-1} = y_0, \dots, X_{n(t)-1+m} = y_m | \mathcal{H}_{n(t)-1}) | \mathcal{F}_t] \\ &= E_z[\mathbf{1}_{\{R_t > s\}} P_{X_t}(T_1 > s_1, \dots, T_m > s_m, X_0 = y_0, \dots, X_m = y_m) | \mathcal{F}_t] \\ &= P_z(R_t > s | \mathcal{F}_t) P_{X_t}(T_1 > s_1, \dots, T_m > s_m, X_0 = y_0, \dots, X_m = y_m) \\ &= P_{X_t}(T_0 > s) P_{X_t}(T_1 > s_1, \dots, T_m > s_m, X_0 = y_0, \dots, X_m = y_m) \\ &= P_{X_t}(T_0 > s | X_0 = y_0, \dots, X_m = y_m) P_{X_t}(T_1 > s_1, \dots, T_m > s_m | X_0 = y_0, \dots, X_m = y_m) \\ &\cdot P_{X_t}(X_0 = y_0, \dots, X_m = y_m) \\ &= P_{X_t}(T_0 > s, T_1 > s_1, \dots, T_m > s_m | X_0 = y_0, \dots, X_m = y_m) P_{X_t}(X_0 = y_0, \dots, X_m = y_m) \\ &= P_{X_t}(T_0 > s, T_1 > s_1, \dots, T_m > s_m, X_0 = y_0, \dots, X_m = y_m) \end{aligned}$$

and this calculation shows that the conditional distribution of (R_t, M_t) given \mathcal{F}_t is the P_{X_t} -distribution of (R_0, M_0) .

In order to complete the proof, we only need to show the statement, that was used in the fourth equality above, i.e. that the conditional distribtuion of R_t given \mathcal{F}_t is the P_{X_t} -distribution of T_0 .

2. Continuous-time Markov chains

So we need to show that for all $s \geq 0$ and all $A \in \mathcal{F}_t$ holds

$$P_z(R_t > s, A) = E_z[P_{X_t}(T_0 > s), A]. \quad (2.11)$$

If $A \subseteq \{\omega(\Delta) \leq t\}$, then both sides are just $P_z(A)$ so we can assume $A \subseteq \{\omega(\Delta) > t\}$. Then it suffices to consider sets of the form $\{n(t) - 1 = n, F_n\}$ for some $n \in \mathbb{N}$ where F_n is as in the proof of Theorem 2.17, i.e. F_n has the form $F_n = \{X_k = i_k, T_{k-1} > t_k, k = 1, \dots, n\}$ for some $i_1, \dots, i_n \in E$ and $t_1, \dots, t_n \geq 0$, because these sets (together with the empty set) form a \cap -stable generator of \mathcal{F}_t . Then

$$\begin{aligned} P_z(R_t > s, n(t) - 1 = n, F_n) &= P_z(J_{n(t)} - t > s, n(t) - 1 = n, F_n) \\ &= P_z(F_n, J_n \leq t \leq t + s < J_{n+1}) \\ &= E_z[F_n, J_n \leq t, P_z(t + s - J_n < T_n | X_k, T_{k-1}, k = 1, \dots, n)] \\ &= E_z[F_n, J_n \leq t, e^{-q(X_n)(t+s-J_n)}] \\ &= E_z[F_n, J_n \leq t, e^{-q(i_n)s} P_z(t < J_n + T_n | X_k, T_{k-1}, k = 1, \dots, n)] \\ &= e^{-q(i_n)s} P_z(F_n, J_n \leq t < J_{n+1}) \end{aligned}$$

where this last term is just the right side of the equation (2.11), thus completing the proof. \square

Remark 2.19. Let $P_{ij}(t), i, j \in E \cup \Delta$ be the transition function of $(X(t))_{t \geq 0}$. Then since Δ is an absorbing state, we have for all $t, s \geq 0$ and $i, j \in E$, $P_{ij}(t+s) = \sum_{k \in E \cup \Delta} P_{ik}(s) P_{kj}(t) = \sum_{k \in E} P_{ik}(s) P_{kj}(t)$, which means that $P_{ij}(t)$ restricted to E is also a standard but possibly dishonest transition function.

After explosions, there would have been several ways of continuing the construction, that would have led to a Markov jump process. The one chosen above obviously minimizes the transition function $P_{ij}(t), i, j \in E$ and is therefore called the minimal construction.

Suppose Q is a conservative q -matrix. We describe now, how we map to Q a Markov jump process. Recall that $q(i) = -q(i, i) < \infty$ and define

$$P(i, j) = \begin{cases} \delta_{ij}, & \text{if } q(i) = 0 \\ 0, & \text{if } q(i) > 0, j = i \\ \frac{q(i, j)}{q(i)}, & \text{if } q(i) > 0, j \neq i \end{cases} \quad (2.12)$$

Because Q is conservative, P is a stochastic matrix. With the $q(i) \geq 0$ and the stochastic matrix P , we construct the minimal Markov jump process as above.

Theorem 2.20. Let Q be a stable and conservative q -matrix, $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ be the corresponding Markov jump process, $P_{ij}(t)$ its transition function (restricted to E). Then $P_{ij}(t)$ satisfies the Kolmogorov backward equation

$$P'_{ij}(t) = \sum_{k \in E} q(i, k) P_{kj}(t) \text{ for all } t \geq 0 \text{ and } j \in E. \quad (2.13)$$

In particular $P'(0) = Q$.

2. Continuous-time Markov chains

Proof. Conditioning upon $T_0 = s$, we get with the help of Theorem 2.17 and bearing in mind that Δ is an absorbing state

$$\begin{aligned} P_{ij}(t) &= \delta_{ij} e^{-q(i)t} + \int_0^t q(i) e^{-q(i)s} \sum_{k \neq i} \frac{q(i, k)}{q(i)} P_{kj}(t-s) ds \\ &= e^{-q(i)t} \left[\delta_{ij} + \int_0^t e^{q(i)s} \sum_{k \neq i} q(i, k) P_{kj}(s) ds \right]. \end{aligned}$$

If we show that $f(s) = \sum_{k \neq i} q(i, k) P_{kj}(s)$ is continuous, we get that $P_{ij}(t)$ is differentiable with derivative

$$\begin{aligned} P'_{ij}(t) &= -q(i) e^{-q(i)t} \left[\delta_{ij} + \int_0^t e^{q(i)s} f(s) ds \right] + e^{-q(i)t} e^{q(i)t} f(t) \\ &= -q(i) P_{ij}(t) + \sum_{k \neq i} q(i, k) P_{kj}(t) = \sum_{k \in E} q(i, k) P_{kj}(t). \end{aligned}$$

If A denotes a finite subset of $E \setminus \{i\}$, then $\sum_{k \in A} q(i, k) P_{kj}(t)$ is continuous and converges uniformly in t to $f(t)$, because

$$\left| f(t) - \sum_{k \in A} q(i, k) P_{kj}(t) \right| \leq \left| \sum_{k \in E \setminus A, k \neq i} q(i, k) P_{kj}(t) \right| \leq \sum_{k \in E \setminus A, k \neq i} q(i, k) \rightarrow 0$$

as $A \nearrow E \setminus \{i\}$, since $\sum_{k \neq i} q(i, k) \leq q(i) < \infty$. Because uniform limits of continuous functions are continuous, $f(t)$ is continuous. \square

Remark 2.21. If $(X(t))$ is a Markov process with given stable and conservative q -matrix Q and if we further assume, that the minimal process of Q is honest, then the transition function $P(t)$ of $(X(t))$ is the only Q -function and thus $(X(t))$ equals the minimal process with probability one. This process can then be constructed as a Markov jump process.

2.4. Classification of states and stationary measures

The following classification of states applies only to Markov jump processes.

Definition 2.22. Given $i, j \in E$, we say that j **can be reached from** i , and write $i \hookrightarrow j$, if $P_{ij}(t) > 0$ for some $t > 0$. We say that i and j **communicate**, if they can be reached from each other and write $i \leftrightarrow j$.

It can be easily shown that \leftrightarrow is an equivalence relation on E and the equivalence classes we call **irreducible classes**. If there is only one irreducible class, we call the Markov process irreducible.

2. Continuous-time Markov chains

Lemma 2.23. *Let $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ be a Markov jump process and $C \subseteq E$. Then C is an irreducible class for $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ if and only if C is an irreducible class for the embedded Markov chain $(X_n)_{n \in \mathbb{N}_0}$.*

Proof. At first, we define the hitting time of $i \in E$ by $T_i = \inf\{t > 0 : X(t) = i\}$. For $j \neq i$ and $t > 0$, we have the equivalence $P_{ij}(t) > 0$ if and only if $P_i(T_j \leq t) > 0$. Indeed we have

$$\begin{aligned} P_{ij}(t) &= E_i[P_i(X(t) = j | T_j)] \\ &\geq E_i[P_i(X(u) = j \text{ for all } u \in [T_j, T_j + t] | T_j) \mathbf{1}_{T_j \leq t}] \\ &= E_i[P_{X(T_j)}(X(u) = j \text{ for all } u \in [0, t]) \mathbf{1}_{T_j \leq t}] \\ &= P_j(T_0 > t) P_i(T_j \leq t) = e^{-q(j)t} P_i(T_j \leq t) > 0, \end{aligned}$$

where we applied the strong Markov property in the third equation.

So let now C be an irreducible class for the embedded Markov chain $(X_n)_{n \in \mathbb{N}_0}$ and $i, j \in C, i \neq j$. That means there exists $n \in \mathbb{N}_0$ and different states $x_1, \dots, x_n \in E$ with $x_i \notin \{i, j\}$ with $P_i(X_1 = x_1, \dots, X_n = x_n, X_{n+1} = j) > 0$. And then for any $t > 0$, we have

$$\begin{aligned} P_i(T_j \leq t) &\geq \\ P_i(T_0 + \dots + T_n \leq t | X_1 = x_1, \dots, X_{n+1} = j) P_i(X_1 = x_1, \dots, X_n = x_n, X_{n+1} = j) &> 0, \end{aligned}$$

because on $\{X_1 = x_1, \dots, X_{n+1} = j\}$ the holding times T_0, \dots, T_n are independent with T_i being exponential with parameter $q(x_i) \neq 0$. So $P_{ij}(t) > 0$ for all $t > 0$.

The other direction works similarly. \square

Definition and Theorem 2.24. *Let $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ be a Markov jump process and $i \in E$. We call the state i **recurrent**, if one of the following equivalent conditions hold:*

1. *The set $\{t : X(t) = i\}$ is unbounded P_i a.s.*
2. *i is recurrent for the embedded Markov chain $(X_n)_{n \in \mathbb{N}_0}$.*

Proof. The proof works similarly but easier as in the previous lemma. \square

Then it's clear from Lemma 2.23 and the corresponding theorem for discrete time Markov chains, that being recurrent is a class property, i.e. if one state of an irreducible class is recurrent, then all states of that class are recurrent. We call a Markov process recurrent, if all states are recurrent.

Definition 2.25. *A measure $\nu \neq 0$ is **stationary** if $0 \leq \nu(j) < \infty$, and $\nu P(t) = \nu$ for all $t > 0$.*

Theorem 2.26. *Suppose that $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ is irreducible and recurrent on E . Then there exists an up to a multiplicative factor unique stationary measure ν . This ν has the property that $\nu(j) > 0$ for all $j \in E$ and can be found in either of the following ways:*

2. Continuous-time Markov chains

(i) For some fixed but arbitrary state i , $v(j)$ is the expected time spent in j between successive entrances to i . That is

$$v(j) = E_i \int_0^{\omega(i)} \mathbf{1}_{\{X(t)=j\}} dt, \quad (2.14)$$

where $\omega(i) = \inf\{t > 0 : X(t) = i, X(t-) \neq i\}$.

(ii) $v(j) = \mu(j)/q(j)$, where μ is stationary for the embedded Markov chain $(X_n)_{n \in \mathbb{N}_0}$.

(iii) As solution to $vQ = 0$.

Proof. We show uniqueness by looking at the embedded Markov chain $(X_n)_{n \in \mathbb{N}_0}$. Because of the last two Theorems, this embedded Markov chain is irreducible and recurrent and it is well known, that an irreducible and recurrent discrete time Markov chain has an up to a multiplicative factor unique stationary measure v that for the irreducibility must satisfy $v(j) > 0$ for all $j \in E$. Since every stationary measure for $(X_t)_{t \geq 0}$ is also a stationary measure for $(X_n)_{n \in \mathbb{N}_0}$ follows the uniqueness.

The proof of (i), we omit here because it works in total analogy of the proof of Theorem 4.13.

For (ii), let $\tau(i) = \inf\{n : X_n = i\}$. Then

$$\begin{aligned} v(j) &= E_i \int_0^{\omega(i)} \mathbf{1}_{\{X(t)=j\}} dt = E_i \left[\sum_{n=0}^{\tau(i)-1} T_n \mathbf{1}_{\{X_n=j\}} \right] \\ &= \sum_{n=0}^{\infty} E_i [T_n, X_n = j, \tau(i) > n] = \sum_{n=0}^{\infty} E_i [E_i [T_n, X_n = j, \tau(i) > n | (X_n)_{n \in \mathbb{N}_0}]] \\ &= \sum_{n=0}^{\infty} E_i [X_n = j, \tau(i) > n, E_i [T_n | (X_n)_{n \in \mathbb{N}_0}]] = \sum_{n=0}^{\infty} E_i \left[X_n = j, \tau(i) > n, \frac{1}{q(X_n)} \right] \\ &= \frac{1}{q(j)} E_i \left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j, \tau(i) > n\}} \right] = \frac{\mu(j)}{q(j)} \end{aligned}$$

where we used in the last step the known result about discrete time Markov chains that $\mu(j) = E_i [\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=j, \tau(i) > n\}}]$ is a stationary measure for any fixed recurrent state $i \in E$.

For (iii), after (ii) v is stationary for $(X(t))_{t \geq 0}$ iff $(v(j)q(j))_{j \in E}$ is stationary for $(X_n)_{n \in \mathbb{N}_0}$, i.e. iff $\sum_{i \in E} v(i)q(i)P(i, j) = v(j)q(j)$ for all $j \in E$. Or since $P(i, i) = 0$ (irreducible implies no state is absorbing) and for $i \neq j$, $q(i, j) = P(i, j)q(i)$ (see (2.12)) iff

$$0 = -v(j)q(j) + \sum_{i \in E, i \neq j} v(i)q(i, j) = \sum_{i \in E} v(i)q(i, j).$$

Lastly, $0 < v(j)$ follows by (ii) since $0 < \mu(j)$. \square

2.5. Examples of continuous-time Markov chains

1. Poisson-Process

Probably the most prominent example of a continuous-time Markov jump process is a Poisson-Process with intensity $\lambda > 0$, i.e. a stochastic process $(N_t)_{t \geq 0}$ with state space \mathbb{N}_0 with the following properties:

1. $N_0 = 0$
2. All sample paths of $(N_t)_{t \geq 0}$ are non-decreasing and right-continuous.
3. For all $s > t \geq 0$ the increment $N_s - N_t$ is Poisson-distributed with parameter $\lambda(s - t)$.
4. $(N_t)_{t \geq 0}$ has independent increments.

However, since Poisson-Processes are of no interest with respect to quasi-stationary distributions, we will not discuss this example any further.

2. Birth-and-Death-Process

Definition 2.27. Let $(\lambda_n)_{n \in \mathbb{N}_0}$ and $(\mu_n)_{n \in \mathbb{N}_0}$ be sequences of non-negative numbers. A continuous-time Markov chain $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ having state space \mathbb{N}_0 and q -matrix given by

$$q(i, j) = \begin{cases} \lambda_i & \text{if } j = i + 1, i \geq 0 \\ \mu_i & \text{if } j = i - 1, i \geq 1 \\ -(\lambda_i + \mu_i) & \text{if } j = i, i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

is called a **birth and death process** on \mathbb{N}_0 with **birth coefficients** λ_n and **death coefficients** $\mu_n, n \geq 0$.

The q -matrix Q of a birth and death process is conservative if and only if $\mu_0 = 0$ and in this case it can be constructed as a nice Markov jump process.

3. Markov branching processes

Definition 2.28. Let $(p_k)_{k \in \mathbb{N}_0}$ denote a sequence of non-negative numbers such that $\sum_{k=0}^{\infty} p_k = 1$, and define

$$p(s) = \sum_{k=0}^{\infty} p_k s^k, \quad 0 \leq s \leq 1.$$

A **continuous-time parameter Markov branching process** with basic generating function $p(s)$ and parameter $a > 0$ is a continuous-time parameter Markov chain with state space \mathbb{N}_0 and q -matrix defined by

$$q(i, j) = \begin{cases} 0 & \text{if } j < i - 1 \\ -ia(1 - p_1) & \text{if } j = i \\ iap_{j-i+1} & \text{if } j \geq i - 1, j \neq i. \end{cases}$$

3. Renewal theory

3.1. Definitions

Definition 3.1. Let τ_1, τ_2, \dots be a sequence of strictly positive independent and identically distributed random variables such that $0 < \mu = E[\tau_1] < \infty$. We refer to τ_i as the *i-th interarrival time*. Define for $n \geq 1$

$$\sigma_n = \sum_{i=1}^n \tau_i.$$

Then σ_n is called the time of the n -th renewal and $(\sigma_n)_{n \in \mathbb{N}_0}$ with the convention $\sigma_0 = 0$ is called a **renewal process**.

We call the common distribution F of τ_1, τ_2, \dots the **interarrival distribution**. Without reference to any renewal process, we call the distribution of a strictly positive random variable with finite expectation an interarrival distribution.

The renewal process is called **periodic**, if there exist $d \in \mathbb{N}$ such that $F(\{nd : n \in \mathbb{N}\}) = 1$ and **aperiodic** otherwise. Define

$$N(t) = \sum_{n=1}^{\infty} \mathbf{1}_{\{\sigma_n \leq t\}}$$

to be the number of renewals up to time t and define

$$R_t = \sigma_{N(t)+1} - t.$$

R_t is called the **remaining life at time t** .

3.2. Iterates of the limit distribution of the remaining life time

Theorem 3.2. Suppose that the renewal process is aperiodic, then

$$\lim_{t \rightarrow \infty} P(R_t > x) = \frac{1}{\mu} \int_x^{\infty} 1 - F(y) dy. \quad (3.1)$$

Proof. See Asmussen [3] Theorem 4.3. on page 155. □

Denote by A the set of aperiodic interarrival distributions. We introduce the map $\Psi : A \rightarrow A$, $F \mapsto \Psi F$ where

$$1 - \Psi F(x) = \lim_{t \rightarrow \infty} P(R_t > x) = \frac{1}{\mu} \int_x^{\infty} 1 - F(y) dy. \quad (3.2)$$

Let now F be an aperiodic interarrival distribution. Assume further that all moments of F exists, that is $m_k(F) = k$ -th moment of $F < \infty$ for all $k \geq 1$. Define $F_k = \Psi^k F$.

3. Renewal theory

These iterates were analysed by Harkness and Shantaram [12]. They already derived the following equations, that we use heavily.

$$1 - F_{k+1}(s) = \frac{1}{m_1(F_k)} \int_s^\infty (1 - F_k(\omega)) d\omega. \quad (3.3)$$

$$m_1(F_k) = \frac{m_{k+1}(F)}{(k+1)m_k(F)}. \quad (3.4)$$

$$\frac{m_k(F_n)}{k!} = \prod_{i=n}^{k+n-1} m_1(F_i). \quad (3.5)$$

3.3. Some convergence results

Theorem 3.3 (Standard method of moments). *Assume there is a unique distribution function F with moments $m_k(F) < \infty, k \geq 1$. If there is a sequence F_n of distribution functions with all moments finite, i.e. $m_k(F_n) < \infty$ for all $n, k \geq 1$, and if*

$$m_k(F_n) \rightarrow m_k(F) \text{ as } n \rightarrow \infty, \text{ for all } k \geq 1.$$

Then F_n converges to F .

Proof. See Chung [7] Theorem 4.5.5. on page 103. □

Theorem 3.4. *Let μ be a measure on \mathbb{R} such that all its moments*

$$m_k = \int_{\mathbb{R}} x^k \mu(dx), k = 0, 1, 2, \dots$$

are finite. If

$$\sum_{k=1}^{\infty} \frac{1}{m_{\frac{1}{2k}}} = \infty, \quad (3.6)$$

then μ is the only measure on \mathbb{R} with $(m_k)_{k \in \mathbb{N}_0}$ as its sequence of moments. (3.6) is called Carleman's condition.

Proof. See Akhiezer [1] on page 85. □

Lemma 3.5. *Assume there exists a subsequence $(n_j)_{j \in \mathbb{N}}$ such that for all $k \geq 1$*

$$\lim_{j \rightarrow \infty} \frac{m_{k+n_j}(F)}{n_j m_{k+n_j-1}(F)} = \theta > 0.$$

Then F_{n_j} converges to an exponential distribution with mean θ .

3. Renewal theory

Proof.

$$\begin{aligned}
m_k(F_{n_j}) &= k! \cdot \prod_{i=n_j}^{k+n_j-1} m_1(F_i) = k! \cdot \prod_{i=n_j}^{k+n_j-1} \frac{m_{i+1}(F)}{(i+1)m_i(F)} \\
&= k! \frac{1}{(n_j+1) \cdots (k+n_j)} \frac{m_{k+n_j}(F)}{m_{n_j}(F)} = k! \prod_{i=1}^k \frac{m_{i+n_j}(F)}{(n_j+i)m_{i+n_j-1}(F)} \\
&= k! \prod_{i=1}^k \frac{m_{i+n_j}(F)}{n_j m_{i+n_j-1}(F)} \cdot \prod_{i=1}^k \frac{n_j}{n_j+i} \xrightarrow{j \rightarrow \infty} k! \theta^k \cdot 1 = k! \theta^k.
\end{aligned}$$

Of course $m_k(\text{Exp}(\theta)) = k! \theta^k$. And since

$$\sum_{k=1}^{\infty} \frac{1}{((2k)! \theta^{2k})^{\frac{1}{2k}}} \geq \frac{1}{\theta} \sum_{k=1}^{\infty} \frac{1}{2k} = \infty.$$

it follows by Carleman's condition, that $\text{Exp}(\theta)$ is the only distribution with all these moments. Then the result follows with the standard method of moments. \square

Lemma 3.6. *Let*

$$\theta = \liminf_{j \rightarrow \infty} \frac{m_j(F)}{j m_{j-1}(F)}.$$

Then there is a subsequence $(j_l)_{l \in \mathbb{N}}$ such that F_{j_l} converges to an exponential distribution with mean θ (where an exponential distribution with mean 0 is δ_0).

Proof. Let $(j_l)_{l \in \mathbb{N}}$ be a subsequence such that

$$\lim_{l \rightarrow \infty} \frac{m_{j_l}(F)}{j_l m_{j_l-1}(F)} = \theta. \tag{3.7}$$

If $\theta = 0$, then $m_1(F_{j_l-1}) = \frac{m_{j_l}(F)}{j_l m_{j_l-1}(F)} \rightarrow 0$, so that F_{j_l} converges to δ_0 and we are done.

So assume $\theta > 0$. $\frac{m_{j+1}(F)}{m_j(F)}$ is increasing in j . Indeed, if τ has distribution F , because of Schwarz-inequality we have

$$\begin{aligned}
(m_j(F))^2 &= (E[\tau^j])^2 = \left(E\left[\tau^{\frac{j-1}{2}} \tau^{\frac{j+1}{2}}\right]\right)^2 \\
&\leq E[\tau^{j-1}] \cdot E[\tau^{j+1}] = m_{j-1}(F) m_{j+1}(F).
\end{aligned}$$

Therefore, for any $k \geq 1$,

$$\begin{aligned}
\theta &= \liminf_{l \rightarrow \infty} \frac{m_{j_l-k}(F)}{(j_l-k) m_{j_l-k-1}(F)} \leq \limsup_{l \rightarrow \infty} \frac{m_{j_l-k}(F)}{(j_l-k) m_{j_l-k-1}(F)} \\
&= \limsup_{l \rightarrow \infty} \frac{m_{j_l-k}(F)}{j_l m_{j_l-k-1}(F)} \leq \limsup_{l \rightarrow \infty} \frac{m_{j_l}(F)}{j_l m_{j_l-1}(F)} = \theta.
\end{aligned}$$

3. Renewal theory

Therefore for each $k \geq 1$,

$$\lim_{l \rightarrow \infty} \frac{m_{j_l - k}(F)}{(j_l - k) m_{j_l - k - 1}(F)} = \theta. \quad (3.8)$$

We are now inductively defining a sub-subsequence $(i_l)_{l \in \mathbb{N}}$ such that for the subsequence $n_l = j_{i_l} - l$ holds for all $k \geq 1$,

$$\lim_{l \rightarrow \infty} \frac{m_{n_l + k}(F)}{(n_l + k) m_{n_l + k - 1}(F)} = \theta. \quad (3.9)$$

If this is shown, then the claim follows with Lemma 3.5.

For shorter notation we define $m_n = \frac{m_n(F)}{n m_{n-1}(F)}$.

Then $m_{j_l - 1}$ converges to θ . So there exists $L_1 > 1$ such that for all $l \geq L_1$ holds $|m_{j_l - 1} - \theta| < 1$. Define $i_1 = L_1$.

$m_{j_l - k}, k = 1, 2$ converge to θ . So there exists $L_2 > L_1 + 1$ such that for all $l \geq L_2$ and $k = 0, 1$ holds $|m_{j_l - 2 + k} - \theta| < \frac{1}{2}$. Define $i_2 = L_2$.

And so on ...

In the n -th step, $m_{j_l - k}, k = 1, 2, \dots, n$ converge to θ . So there exists $L_n > L_{n-1} + 1$ such that for all $l \geq L_n$ and $0 \leq k < n$ holds $|m_{j_l - n + k} - \theta| < \frac{1}{n}$. Then we define $i_n = L_n$.

Then for any fixed $k \geq 1$, $m_{n_l + k}$ converges to θ , because for any $l > k$, $|m_{n_l + k} - \theta| < \frac{1}{l}$, which finishes the proof. □

Theorem 3.7. *If for some $C < \infty, \lambda > 0$, it holds that $1 - F(t) \leq Ce^{-\lambda t}$ for all $t \geq 0$, then there exists a $\theta \leq \frac{1}{\lambda}$ and a subsequence $(n_j)_{j \in \mathbb{N}}$ such that F_{n_j} converges to an exponential distribution with mean θ as $j \rightarrow \infty$.*

Proof. We apply the formula of integration by parts of the Stieltjes-integral:

$$\begin{aligned} m_k(F) &= \int_0^\infty t^k dF(t) = \int_0^\infty -t^k d(1 - F(t)) \\ &= [-t^k(1 - F(t))]_0^\infty + k \int_0^\infty t^{k-1}(1 - F(t)) dt = k \int_0^\infty t^{k-1}(1 - F(t)) dt \\ &\leq Ck \int_0^\infty t^{k-1} e^{-\lambda t} dt = Ck! \lambda^{-k}. \end{aligned}$$

Thus

$$C\lambda^{-k} \geq \frac{m_k(F)}{k!} = \prod_{j=0}^{k-1} m_1(F_j) = \prod_{j=1}^k \frac{m_j(F)}{j m_{j-1}(F)}.$$

It follows

3. Renewal theory

$$\theta := \liminf_{j \rightarrow \infty} \frac{m_j(F)}{jm_{j-1}(F)} \leq \frac{1}{\lambda} \tag{3.10}$$

and we only need to apply the previous Lemma to get the result. □

4. The Main Theorem

4.1. Setting and Theorem

Our basic setting for the rest of this paper is the following:

We consider $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ a continuous time Markov chain with state space \mathbb{N}_0 with a given q -matrix Q . We assume Q to be stable and conservative. Further, we assume that the minimal process for Q is honest. Then we may assume that $(X(t))_{t \in \mathbb{R}_{\geq 0}}$ is a nice Markov jump process because of Remark 2.21. In particular,

$$P_\mu(X(t) = \Delta) = 0 \text{ for all } t \geq 0 \text{ and all initial distributions } \mu \text{ on } \mathbb{N}.$$

Further we assume 0 to be an absorbing state, i.e. $q(0) = 0$ and that all other states form an irreducible class. We define

$$R := \inf\{t \geq 0 : X(t) = 0\}$$

the **absorption, killing or extinction time** at 0 and we assume that absorption is certain, i.e. $P_x(R < \infty) = 1$ for all $x \geq 1$. This implies obviously that there is no explosion.

For a probability distribution μ on \mathbb{N} , we further define F^μ as the distribution of the absorption time of the process starting out with initial distribution μ , i.e.

$$F^\mu = P_\mu(R \in \bullet). \quad (4.1)$$

Definition 4.1. A probability measure μ on \mathbb{N} (also seen as a probability distribution on $\mathbb{N} \cup \{0\}$ with $\mu(0) = 0$) is called a **quasi-stationary distribution** if, starting with μ , the conditional distribution, given the event the process hasn't been absorbed, is still μ , i.e.

$$P_\mu(X(t) = \bullet | X(t) \neq 0) = P_\mu(X(t) = \bullet | R > t) = \mu.$$

We see immediately, that the irreducibility condition implies that every quasi-stationary distribution μ charges all points in \mathbb{N} .

As we already motivated in the introduction, every quasi-limiting distribution (or Yaglom-limit) must be a quasi-stationary distribution. For this, we give now a detailed proof.

Definition 4.2. A probability measure π on \mathbb{N} is called a **Yaglom-limit** if for all $x, y \in \mathbb{N}$ holds

$$\pi(y) = \lim_{t \rightarrow \infty} P_x(X(t) = y | R > t).$$

Lemma 4.3. Assume that π is a Yaglom-limit. Then for all bounded functions $f : \mathbb{N} \rightarrow \mathbb{R}$ and all $x \in \mathbb{N}$ holds

$$E_\pi[f] = \sum_{y \in \mathbb{N}} \pi(y) f(y) = \lim_{t \rightarrow \infty} E_x[f(X(t)) | R > t].$$

4. The Main Theorem

Proof. So let $x \in \mathbb{N}$ be fixed.

Obviously for all finite $A \subseteq \mathbb{N}$ holds $\pi(A) = \lim_{t \rightarrow \infty} P_x(X(t) \in A | R > t)$ and the function

$$t \mapsto P_x(X(t) \in A | R > t) = \frac{P_x(X(t) \in A \cap \mathbb{N})}{P_x(X(t) \in \mathbb{N})} = \frac{\sum_{y \in A \cap \mathbb{N}} P_{xy}(t)}{1 - P_{x0}(t)}$$

is continuous. From here, we need the following result.

Lemma 4.4. *The family $(P_x(X(t) \in \bullet | R > t))_{t \geq 0}$ is **tight**, i.e. for all $\epsilon > 0$ there exists $M_\epsilon \in \mathbb{N}$ such that*

$$P_x(X(t) \in [1, M_\epsilon] | R > t) \geq 1 - \epsilon \quad \text{for all } t \geq 0.$$

Proof. So let $\epsilon > 0$ be arbitrary. There is M_ϵ^* such that $\pi([1, M_\epsilon^*]) > 1 - \epsilon$. Because

$$\lim_{t \rightarrow \infty} P_x(X(t) \in [1, M_\epsilon^*] | R > t) = \pi([1, M_\epsilon^*]) > 1 - \epsilon,$$

it follows that there is some $T > 0$ such that for all $t > T$ we have $P_x(X(t) \in [1, M_\epsilon^*] | R > t) > 1 - \epsilon$.

For all $t \in [0, T]$ there exists $M_{\epsilon, t} \in \mathbb{N}$ such that $P_x(X(t) \in [1, M_{\epsilon, t}] | R > t) > 1 - \frac{\epsilon}{2}$. Because for all $t \in [0, T]$ the function $s \mapsto P_x(X(s) \in [1, M_{\epsilon, t}] | R > s)$ is continuous, there is $\delta_t > 0$ such that for all $s \in K_{\delta_t}(t)$ holds:

$$|P_x(X(s) \in [1, M_{\epsilon, t}] | R > s) - P_x(X(t) \in [1, M_{\epsilon, t}] | R > t)| \leq \frac{\epsilon}{2}.$$

Since $[0, T] \subseteq \cup_{t \in [0, T]} K_{\delta_t}(t)$ is an open cover of the compact set $[0, T]$ there exists $M \in \mathbb{N}$ and $t_1, \dots, t_M \in [0, T]$ such that $[0, T] \subseteq \cup_{i=1}^M K_{\delta_{t_i}}(t_i)$. Finally if we choose $M_\epsilon = \max\{M_\epsilon^*, M_{\epsilon, t_1}, \dots, M_{\epsilon, t_M}\}$ for all $t \geq 0$ it holds

$$P_x(X(t) \in [1, M_\epsilon] | R > t) \geq 1 - \epsilon.$$

□

Now if $A \subseteq \mathbb{N}$ (not necessarily finite) with the just proven tightness, we can show that $\pi(A) = \lim_{t \rightarrow \infty} P_x(X(t) \in A | R > t)$. This argument is omitted here, because the same idea is presented later in the proof.

Then obviously for all step functions $f : \mathbb{N} \rightarrow \mathbb{R}$, i.e. for all f of the form $f = \sum_{i=1}^m c_i \mathbf{1}_{A_i}$ with $m \in \mathbb{N}$, $c_1, \dots, c_m \in \mathbb{R}$ and $A_1, \dots, A_m \subseteq \mathbb{N}$ disjoint holds $E_\pi[f] = \lim_{t \rightarrow \infty} E_x[f(X(t)) | R > t]$. For the last step, let $f : \mathbb{N} \rightarrow \mathbb{R}$ be bounded. Then the sequence of functions $(f_n)_{n \in \mathbb{N}}$ given by

$$f_n = f \mathbf{1}_{[1, n]} = \sum_{y=1}^n f(y) \mathbf{1}_{\{y\}}$$

converges to f . So we have by bounded convergence $\lim_{t \rightarrow \infty} E_x[f(X(t)) | R > t] = \lim_{t \rightarrow \infty} \lim_{n \rightarrow \infty} E_x[f_n(X(t)) | R > t]$. Now in order to exchange the limits above, we

4. The Main Theorem

show that the limit $\lim_{n \rightarrow \infty} E_x [f_n(X(t)) | R > t] = E_x [f(X(t)) | R > t]$ is reached uniformly in t .

So let $\epsilon > 0$ be arbitrary. There is $M_\epsilon \in \mathbb{N}$ such that for all $t \geq 0$ holds:

$$P_x(X(t) \in [1, M_\epsilon] | R > t) \geq 1 - \epsilon.$$

Then for all $n \geq M_\epsilon$ holds:

$$\begin{aligned} & |E_x [f(X(t)) | R > t] - E_x [f_n(X(t)) | R > t]| \\ & \leq E_x [|f(X(t)) - f_n(X(t))| \mathbf{1}_{\{X(t) \in [1, M_\epsilon]\}} | R > t] + 2 \|f\|_\infty P_x(X(t) \in [M_\epsilon + 1, \dots, \infty] | R > t) \\ & \leq 2 \|f\|_\infty \epsilon. \end{aligned}$$

So we can exchange the limits and continuing the calculation, we get

$$\lim_{t \rightarrow \infty} E_x [f(X(t)) | R > t] = \lim_{n \rightarrow \infty} \lim_{t \rightarrow \infty} E_x [f_n(X(t)) | R > t] = \lim_{n \rightarrow \infty} E_\pi [f_n] = E_\pi [f].$$

□

Theorem 4.5. *Assume that π is a Yaglom-limit. Then π is also a quasi-stationary distribution.*

Proof. By Lemma 4.3 there exists a probability measure μ on \mathbb{N} (a Dirac measure) such that for all bounded functions $f : \mathbb{N} \rightarrow \mathbb{R}$ it holds that

$$E_\pi [f] = \sum_{x \in \mathbb{N}} \pi(x) f(x) = \lim_{t \rightarrow \infty} E_\mu (f(X_t) | R > t).$$

Applying this to the bounded function $f(z) = P_z(R > s)$ we get with the Markov property and the definition $t + R = \inf \{s \geq t : X(s) = 0\}$

$$\begin{aligned} \lim_{t \rightarrow \infty} E_\mu (f(X_t) | R > t) &= \lim_{t \rightarrow \infty} \frac{E_\mu [P_{X_t}(R > s) \mathbf{1}_{\{R > t\}}]}{P_\mu(R > t)} \\ &= \lim_{t \rightarrow \infty} \frac{E_\mu [P_\mu(t + R > s + t | \mathcal{F}_t) \mathbf{1}_{\{R > t\}}]}{P_\mu(R > t)} \\ &= \lim_{t \rightarrow \infty} \frac{P_\mu(R > s + t)}{P_\mu(R > t)} \\ &= P_\pi(R > s). \end{aligned}$$

Now let us denote by $g(s) = P_\pi(R > s)$ this last quantity. We can show that g is bounded, monotone decreasing, non-zero and satisfies the Cauchy exponential functional

4. The Main Theorem

equation $g(s + s') = g(s)g(s')$. Indeed with the definition $f_{s'}(z) = P_z(R > s')$ we have

$$\begin{aligned}
g(s + s') &= P_\pi(R > s + s') = \lim_{t \rightarrow \infty} \frac{P_\mu(R > t + s + s')}{P_\mu(R > t)} \\
&= \lim_{t \rightarrow \infty} \frac{E_\mu \left[P_\mu(t + s + R > t + s + s' | \mathcal{F}_{t+s}) \mathbf{1}_{\{R > t+s\}} \right]}{P_\mu(R > t)} \\
&= \lim_{t \rightarrow \infty} \frac{E_\mu \left[P_{X(t+s)}(R > s') \mathbf{1}_{\{R > t+s\}} \right]}{P_\mu(R > t+s)} \frac{P_\mu(R > t+s)}{P_\mu(R > t)} \\
&= \lim_{t \rightarrow \infty} E_\mu[f_{s'}(X(t+s)) | R > t+s] \frac{P_\mu(R > t+s)}{P_\mu(R > t)} \\
&= P_\pi(R > s') \lim_{t \rightarrow \infty} \frac{P_\mu(R > t+s)}{P_\mu(R > t)} \\
&= P_\pi(R > s') P_\pi(R > s) = g(s')g(s).
\end{aligned}$$

Thus, there exists $0 < a < \infty$ such that $g(s) = P_\pi(R > s) = e^{-as}$. We now consider for any $y \in \mathbb{N}$ the bounded function $f(z) = P_z(X(s) = y, R > s)$. By the Markov property, we get

$$\begin{aligned}
P_\pi(X(s) = y | R > s) P_\pi(R > s) &= P_\pi(X(s) = y, R > s) = E_\pi[f] \\
&= \lim_{t \rightarrow \infty} E_\mu[P_{X(t)}(X(s) = y, R > s) | R > t] \\
&= \lim_{t \rightarrow \infty} \frac{E_\mu[P_{X(t)}(X(s) = y, R > s) \mathbf{1}_{\{R > t\}}]}{P_\mu(R > t)} \\
&= \lim_{t \rightarrow \infty} \frac{E_\mu[P_\mu(X(s+t) = y, t+R > s+t | \mathcal{F}_t) \mathbf{1}_{\{R > t\}}]}{P_\mu(R > t)} \\
&= \lim_{t \rightarrow \infty} \frac{P_\mu(X(s+t) = y, R > s+t)}{P_\mu(R > s+t)} \frac{P_\mu(R > s+t)}{P_\mu(R > t)}
\end{aligned}$$

For $t \rightarrow \infty$, the term $\frac{P_\mu(X(s+t)=y, R>s+t)}{P_\mu(R>s+t)}$ goes to $\pi(y)$ by definition, whereas the second term $\frac{P_\mu(R>s+t)}{P_\mu(R>t)}$ goes to $P_\pi(R > s) = e^{-as} > 0$ as we showed in the first part of the proof. Thus we showed that for any $s > 0, y \in \mathbb{N}$ the equation $\pi(y) = P_\pi(X(s) = y | R > s)$ holds and that π is indeed a quasi-stationary distribution. \square

We immediately state the main result of this paper.

Theorem 4.6. *Assume that*

$$\lim_{x \rightarrow \infty} P_x(R < t) = 0, \quad \forall t \geq 0. \quad (4.2)$$

Then a quasi-stationary distribution exists if and only if

$$E_x[e^{\lambda R}] < \infty \quad (4.3)$$

for some $\lambda > 0$ and for some $x \in \{1, 2, \dots\}$.

4. The Main Theorem

Condition (4.2) of **not coming from infinity in finite time**, is satisfied in many examples (think of a birth-and-death process) and can be checked directly from the transition rates.

If (4.3) holds, we can define the quantity $\lambda_0 = \sup\{\lambda : E_x[e^{\lambda R}] < \infty\} > 0$ and call it the **exponential rate of survival** of the process. This quantity is independent of $x \in \mathbb{N}$, as we see below. In the book of Martinez (page 5 and Theorem 5.4. on page 80), it can be read that for the birth and death case, there is no unique quasi-stationary distribution but a continuum, in fact for each $\lambda \in (0, \lambda_0]$ there is one quasi-stationary distribution μ_λ .

Remark 4.7. *If (4.3) holds for some $x \in \mathbb{N}$, then it holds for all $y \in \mathbb{N}$, because \mathbb{N} was assumed an irreducible class, so if $x \neq y \in \mathbb{N}$, then there exists $s > 0$ such that $P_x(X(s) = y) > 0$. Then*

$$\begin{aligned}
 \infty > E_x[e^{\lambda R}] &= \int_0^\infty P_x(e^{\lambda R} > t) dt \\
 &\geq \int_0^\infty P_x(e^{\lambda R} > t, X(s) = y) dt \\
 &= P_x(X(s) = y) \cdot \int_0^\infty P(e^{\lambda R} > t | X(s) = y) dt \\
 &= P_x(X(s) = y) \cdot \int_0^\infty P\left(X(u) \neq 0 \text{ for } u \in \left[s, \frac{1}{\lambda} \ln(t)\right] | X(s) = y\right) dt \\
 &= P_x(X(s) = y) \cdot \int_0^\infty P_y\left(X(u) \neq 0 \text{ for } u \in \left[0, \frac{1}{\lambda} \ln(t) - s\right]\right) dt \\
 &= P_x(X(s) = y) \cdot \int_0^\infty P_y(e^{\lambda R} > te^{-\lambda s}) dt \\
 &\geq P_x(X(s) = y) \cdot \int_0^\infty P_y(e^{\lambda R} > t) dt \\
 &= P_x(X(s) = y) \cdot E_y[e^{\lambda R}]
 \end{aligned}$$

and thus also $E_y[e^{\lambda R}] < \infty$.

Remark 4.8. *All $\lambda > 0$ satisfying (4.3) must verify*

$$\lambda < q(x) \quad \text{for all } x \in \mathbb{N}.$$

Proof. So fix $x \in \mathbb{N}$ and assume that $\lambda > 0$ is such that $E_x[e^{\lambda R}] < \infty$. Then since for the time of the first jump J_1 holds $P_x(J_1 \leq R) = 1$, we get $E_x[e^{\lambda J_1}] < \infty$. But since for $q(x) \leq \lambda$ holds $E_x[e^{\lambda J_1}] = \int_0^\infty e^{\lambda y} q(x) e^{-q(x)y} dy = \infty$, we can deduce that $q(x) > \lambda$. \square

Remark 4.9. *Condition (4.3) is equivalent to exponential decay of $P_x(R \in \bullet)$, i.e.*

$$\text{there exist } C, \gamma > 0 \text{ such that } P_x(R > t) \leq Ce^{-\gamma t} \quad \text{for all } t \geq 0. \quad (4.4)$$

4. The Main Theorem

Proof. If $E_x[e^{\lambda R}] < \infty$ for some $\lambda > 0$ and some $x \in \mathbb{N}$. Then by Markov inequality $P_x(R > t) = P_x(e^{\lambda R} > e^{\lambda t}) \leq E_x[e^{\lambda R}] e^{-\lambda t}$ and the claim follows with $\gamma = \lambda$ and $C = E_x[e^{\lambda R}]$.

On the other hand, if (4.4) holds. Then let $\lambda < \gamma$ and then $E_x[e^{\lambda R}] = 1 + E_x[e^{\lambda R} - 1] = 1 + \int_0^\infty P_x(e^{\lambda R} - 1 > t) dt = 1 + \int_0^\infty P_x(R > \frac{1}{\lambda} \ln(t+1)) dt \leq 1 + \int_0^\infty C e^{-\gamma \frac{1}{\lambda} \ln(t+1)} dt = 1 + C \int_1^\infty e^{-\gamma \frac{1}{\lambda} \ln(t)} dt = 1 + C \int_1^\infty t^{-\frac{\gamma}{\lambda}} dt < \infty$. \square

Proof of Theorem 4.6. We only prove that this is a necessary condition here and sketch the proof for the other direction. So let μ be a quasi-stationary distribution. Then for $s, t \geq 0$ holds

$$\begin{aligned} & P_\mu(R > s+t | R > s) \\ &= P_\mu(X(s+t) \neq 0 | X(s) \neq 0) \\ &= \sum_{y \in \mathbb{N}} P_\mu(X(s+t) \neq 0, X(s) = y | X(s) \neq 0) \\ &= \sum_{y \in \mathbb{N}} P(X(s+t) \neq 0 | X(s) = y) \cdot P_\mu(X(s) = y | X(s) \neq 0) \\ &= \sum_{y \in \mathbb{N}} P_y(X(t) \neq 0) \cdot \mu(y) \\ &= P_\mu(X(t) \neq 0) \\ &= P_\mu(R > t). \end{aligned}$$

Therefore the function $g(t) = P_\mu(R > t)$ defined on $R_{\geq 0}$ must satisfy the Cauchy exponential functional equation $g(t+s) = g(t)g(s)$. Since $g \neq 0$, $g \leq 1$ and g monotone decreasing, $g(t) = e^{Ct}$ for some constant $C \leq 0$. If C would be 0, then $P_\mu(R = \infty) = \lim_{t \rightarrow \infty} P_\mu(R > t) = 1$, but because of our basic setting $P_\mu(R < \infty) = \sum_{x \in \mathbb{N}} \mu(x) P_x(R < \infty) = 1$, so $C < 0$. So $P_\mu(R \in \bullet)$ is exponentially distributed with parameter $-C = \frac{1}{E_\mu[R]}$. Then for all $\lambda < \frac{1}{E_\mu[R]}$ we have that $E_\mu[e^{\lambda R}] < \infty$. Since μ is a quasi-invariant distribution and \mathbb{N} an irreducible class, then $\mu(x) > 0$ must be for all $x \in \mathbb{N}$. Thus from $E_x[e^{\lambda R}] \leq \frac{1}{\mu(x)} E_\mu[e^{\lambda R}] < \infty$ follows the claim. \square

We now sketch as good as possible the proof for the other direction and highlight important intermediate results as a perspective to the following chapters:

1. For any probability measure μ on \mathbb{N} , we have the stable and conservative q -matrix $Q^\mu = (q^\mu(x, y))_{x, y \in \mathbb{N}}$ given by

$$q^\mu(x, y) = q(x, y) + q(x, 0) \mu(y), x, y \in \mathbb{N}. \quad (4.5)$$

It is the topic of section 4.2 to show that if $E_\mu R < \infty$, the Markov jump process Y^μ as defined in (4.10) is irreducible and recurrent and has Q^μ as its q -matrix.

4. The Main Theorem

With the help of Y^μ we prove that the equation $vQ^\mu = 0$ has a unique probability distribution solution v with $v(y) > 0$ for all $y \in \mathbb{N}$ and even more usefully

$$v(y) = \frac{1}{E_\mu R} \int_0^\infty P_\mu(X(t) = y) dt. \quad (4.6)$$

2. In section 4.3, we prove that a fixpoint of the map

$$\Phi : \mu \mapsto v, \quad E_\mu R < \infty \quad (4.7)$$

is a quasi-stationary distribution for $X(t)$, where we use that $X(t)$ is the minimal process for Q and that absorption is sure. In the next sections, we prove the existence of a fixpoint of Φ .

3. We prove the existence of a fixpoint of Φ by the Schauder-Tychonov fixpoint theorem.

Theorem 4.10 (Schauder-Tychonov fixpoint theorem). *Let X be a Hausdorff, topological vectorspace, $\emptyset \neq K \subseteq X$ convex and compact. Then every continuous map $\Phi : K \rightarrow K$ has at least one fixpoint.*

Proof. See Cauty [6]. □

In our case, X is the vectorspace over \mathbb{R} of all functions from \mathbb{N} to \mathbb{R} , with the following metric defined on it, that makes X a Hausdorff, topological vectorspace:

$$d(\mu, \mu') = \sum_{y \in \mathbb{N}} \frac{1}{2^y} \frac{|\mu(y) - \mu'(y)|}{1 + |\mu(y) - \mu'(y)|}. \quad (4.8)$$

A sequence $(\mu_k)_{k \in \mathbb{N}} \subseteq X$ converges to $\mu \in X$ with respect to this metric, if and only if μ_k converges to μ pointwise. In metric spaces, sequential continuity is the same as continuity and sequential compactness is the same as compactness.

K we choose as \mathcal{M}_θ with

$$\begin{aligned} \mathcal{M}_\theta &= \{\mu : \mu \text{ probability measure on } \mathbb{N} \text{ such that } P_\mu(R > t) = e^{-\frac{t}{\theta}}\} \\ &= \{\mu : \mu \text{ probability measure on } \mathbb{N} \text{ such that } F^\mu = P_\mu(R \in \bullet) \\ &\quad \text{is an exponential distribution with mean } \theta\} \end{aligned} \quad (4.9)$$

for some $\theta > 0$ such that this set is not empty. We note here that for $\mu \in \mathcal{M}_\theta$ we have $E_\mu R = \theta < \infty$, so $\Phi\mu$ is defined.

4. In section 4.4, we check all assumptions from the Schauder-Tychonov theorem, i.e. if $\mathcal{M}_\theta \neq \emptyset$, then we show
- a) $\Phi(\mathcal{M}_\theta) \subseteq \mathcal{M}_\theta$.
 - b) $\Phi : \mathcal{M}_\theta \rightarrow \mathcal{M}_\theta$ is (sequentially) continuous.

4. The Main Theorem

c) \mathcal{M}_θ is convex.

d) \mathcal{M}_θ is (sequentially) compact.

(4a) and (4b) follow easily from (4.6). (4c) is trivial. For (4d), we show that any sequence $(\mu_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}_\theta$ is tight, where we critically use the assumption $\lim_{x \rightarrow \infty} P_x(R < t) = 0$ for all t . Then we use Prohorov's theorem to show sequential compactness.

5. The hard part is to show, that there exists $\theta > 0$ such that $\mathcal{M}_\theta \neq \emptyset$. This is actually the only part, where we use the assumption $E_x e^{\lambda R} < \infty$ for some $\lambda > 0$.

If F^μ denotes the distribution of the absorption time, if $X(t)$ starts out with initial distribution μ , then we show that F^{Φ^μ} is the limit distribution of the remaining life time of a renewal process with interarrival distribution F^μ , see Corollary 4.29. This allows us to apply our convergence results from renewal theory, and prove that if F^{δ_x} has exponentially bounded tails (which is equivalent to $E_x e^{\lambda R} < \infty$ for some $\lambda > 0$), then $F^{\Phi^{n\delta_x}}$ converges along subsequences to an exponential distribution.

4.2. A stationary Markov jump process related to qsd.

For any probability distribution μ on \mathbb{N} , we now construct a Markov jump process $(Y^\mu(t))_{t \in \mathbb{R}_{\geq 0}}$ with q -matrix Q^μ , such that it is related to the process $(X(t))_{t \in \mathbb{R}_{\geq 0}}$.

Let $\{X_k(t) : k = 1, 2, \dots\}$ be a sequence of independent copies of $X(t)$ with initial distribution μ and absorption times $t_k = \inf\{t > 0 : X_k(t) = 0\}$. Then define

$$\begin{aligned} Y^\mu(t) &= X_1(t) && \text{for } 0 \leq t < t_1, \\ Y^\mu(t_1 + t) &= X_2(t) && \text{for } 0 \leq t < t_2, \\ &\dots \end{aligned}$$

Formally, with $s_0 = 0, s_k = \sum_{i=1}^k t_i$ for $k \geq 1$,

$$Y^\mu(t) = \sum_{k=1}^{\infty} X_k(t - s_{k-1}) \mathbf{1}_{\{t \in [s_{k-1}, s_k)\}}. \quad (4.10)$$

Note, that the initial distribution of Y^μ is just the initial distribution of X_1 . For $i \in \mathbb{N}$ define

$$H_{ij}(t) = P_i(t_1 \leq t, Y^\mu(t_1) = j).$$

We have $t < t_1 < t + dt$ if $X_1(t) = k$ for some $k \in \mathbb{N}$, then the absorption occurs with probability $q(k, 0) + o(dt)$ and the new state after resurrection is j with probability $\mu(j)$. So

$$H_{ij}(t) = \int_0^t \sum_{k \in \mathbb{N}} P_{ik}(v) q(k, 0) \mu(j) dv.$$

4. The Main Theorem

So $H_{ij}(t)$ is continuously differentiable and obviously non-decreasing.

By definition of $Y^\mu(t)$, $Y^\mu(t) = j$ if and only if $X_1(t) = j$ or if for $v \in (0, t)$, $t_1 = v$, $Y^\mu(v) = k \in \mathbb{N}$ and Y^μ makes a transition from k to j during the time $t - v$. For the transition function $P_{i,j}^\mu(t)$ of Y^μ holds then

$$P_{i,j}^\mu(t) = P_{ij}(t) + \int_0^t \sum_{k \in \mathbb{N}} H'_{ik}(v) P_{k,j}^\mu(t-v) dv. \quad (4.11)$$

Theorem 4.11. 1. (4.11) has a minimal transition function solution $P_{i,j}^\mu(t)$.

2. Any transition function solution of (4.11) has the q -matrix given by Q^μ .

Proof. (4.11) has the same form as equation (2.1) on page 67 of Anderson [2]. So we just need to check the assumptions in Lemma 2.1. of Anderson. We already mentioned that H_{ij} is non-decreasing and continuously differentiable. Also, because of Fubini's theorem

$$\begin{aligned} H'_{ij}(t+v) &= \sum_{k \in \mathbb{N}} P_{ik}(t+v) q(k, 0) \mu(j) \\ &= \sum_{k \in \mathbb{N}} \sum_{l \in \mathbb{N}} P_{il}(t) P_{lk}(v) q(k, 0) \mu(j) \\ &= \sum_{l \in \mathbb{N}} P_{il}(t) H'_{lj}(v) \end{aligned}$$

and $\sum_{j \in \mathbb{N}} (P_{ij}(t) + H_{ij}(t)) = P_i(t_1 > t) + P_i(t_1 \leq t) = 1$. The claim of 1 follows now with Lemma 2.1. of Anderson.

For 2, let $P_{i,j}^\mu(t)$ be a transition function solution of (4.11). Applying Fubini and the integral mean value theorem to (4.11) yields

$$P_{i,j}^\mu(t) = P_{ij}(t) + \sum_{k \in \mathbb{N}} H_{ik}(t) P_{k,j}^\mu(\xi_{kj}),$$

where $0 < \xi_{kj} < t$. In addition,

$$H'_{ik}(0) = \lim_{t \rightarrow 0} \frac{H_{ik}(t)}{t} = q(i, 0) \mu(k).$$

If Ω_{ij} denotes the q -matrix of $P_{i,j}^\mu(t)$, then we apply Fatous lemma to get

$$\begin{aligned} \Omega_{ij} &= P_{i,j}^{\prime\mu}(0) \geq P'_{ij}(0) + \liminf_{t \rightarrow 0} \frac{\sum_{k \in \mathbb{N}} H_{ik} P_{k,j}^\mu(\xi_{kj})}{t} \\ &\geq q(i, j) + \sum_{k \in \mathbb{N}} \liminf_{t \rightarrow 0} \frac{H_{ik} P_{k,j}^\mu(\xi_{kj})}{t} \geq q(i, j) + \sum_{k \in \mathbb{N}} \liminf_{t \rightarrow 0} \frac{H_{ik}}{t} \cdot \liminf_{t \rightarrow 0} P_{k,j}^\mu(\xi_{kj}) \\ &= q(i, j) + \sum_{k \in \mathbb{N}} q(i, 0) \mu(k) \cdot \delta_{kj} = q(i, j) + q(i, 0) \mu(j). \end{aligned}$$

4. The Main Theorem

So $\sum_{j \in \mathbb{N}} \Omega_{ij} \geq \sum_{j \in \mathbb{N}} q^\mu(i, j) = 0$, because Q^μ is conservative. As $\sum_{j \in \mathbb{N}} \Omega_{ij}$ can't be positive (see Theorem 2.7), $\Omega_{ij} = q^\mu(i, j)$. \square

Lemma 4.12. *If $E_\mu R < \infty$, then $(Y^\mu(t))_{t \in \mathbb{R}_{\geq 0}}$ is irreducible and recurrent.*

Proof. $(Y^\mu(t))_{t \in \mathbb{R}_{\geq 0}}$ is irreducible follows from \mathbb{N} being an irreducible class of $(X(t))_{t \in \mathbb{R}_{\geq 0}}$.

Recurrence is a class property, so we just need to find one recurrent state. Let $i \in \mathbb{N}$ with $\mu(i) > 0$. We prove

$$P_i(\{t : Y^\mu(t) = i\} \text{ unbounded}) = 1,$$

which is equivalent to i being recurrent. $P(\sum_{n \in \mathbb{N}} t_n = \infty) = 1$, because since the t_n are all independent and identically distributed according to F^μ with finite expectation $E_\mu[R] < \infty$, we get from the strong law of large numbers $\frac{s_n}{n} \rightarrow E_\mu[R]$ almost surely and thus $s_n \rightarrow \infty$ almost surely.

Also $P(t_n < \infty \text{ for all } n) = 1$, then

$$P_i(\{t : Y^\mu(t) = i\} \text{ unbounded}) \geq P(Y^\mu(s_n) = i \text{ for infinitely many } n) = 1,$$

where the last equality follows from Borel-Cantelli-lemma since $\{Y^\mu(s_n) = i\}$ are independent events and $\sum_{n \in \mathbb{N}} P(Y^\mu(s_n) = i) = \sum_{n \in \mathbb{N}} \mu(i) = \infty$. \square

Since Y^μ is irreducible and recurrent, by Theorem 2.26, we get an up to a multiplicative factor unique invariant measure v on \mathbb{N} of Y^μ satisfying $vQ^\mu = 0$ with the property that $v(y) > 0$ for all $y \in \mathbb{N}$. In the following theorem, we show that v can be normalized to a probability distribution.

Theorem 4.13. *If $E_\mu R < \infty$, then there exists a unique invariant distribution Φ_μ of Y^μ , given by*

$$\Phi_\mu(y) = \frac{1}{E_\mu R} \int_0^\infty P_\mu(X(t) = y) dt. \quad (4.12)$$

Proof. $\frac{1}{E_\mu R}$ is just a constant, so we first show that v given by $v(j) = \int_0^\infty P_\mu(X(t) = j) dt$ is an invariant measure for Y^μ . We also note immediately that

$$\begin{aligned} E_\mu \int_0^{t_1} \mathbf{1}_{\{Y^\mu(t)=j\}} dt &= E_\mu \int_0^\infty \mathbf{1}_{\{Y^\mu(t)=j, t_1 > t\}} dt = E_\mu \int_0^\infty \mathbf{1}_{\{X_1(t)=j\}} dt \\ &= \int_0^\infty P_\mu(X(t) = j) dt = v(j) \end{aligned}$$

where we applied Fubini's theorem in the next to last step.

4. The Main Theorem

Then we note that $(Y^\mu(t))_{0 \leq t \leq h}$ and $(Y^\mu(t+t_1))_{0 \leq t \leq h}$ have the same P_μ distribution. Thus for all $h > 0$,

$$\begin{aligned} E_\mu \int_0^{t_1} \mathbf{1}_{\{Y^\mu(t)=j\}} dt &= E_\mu \int_0^h + \int_h^{t_1} \mathbf{1}_{\{Y^\mu(t)=j\}} dt = E_\mu \int_{t_1}^{t_1+h} + \int_h^{t_1} \mathbf{1}_{\{Y^\mu(t)=j\}} dt \\ &= E_\mu \int_h^{t_1+h} \mathbf{1}_{\{Y^\mu(t)=j\}} dt = E_\mu \int_0^{t_1} \mathbf{1}_{\{Y^\mu(t+h)=j\}} dt. \end{aligned}$$

And this is also true if $t_1 < h$. So

$$\begin{aligned} E_\mu \int_0^{t_1} \mathbf{1}_{\{Y^\mu(t)=j\}} dt &= \int_0^\infty E_\mu [\mathbf{1}_{\{Y^\mu(t+h)=j, t_1 > t\}}] dt \\ &= \int_0^\infty E_\mu [E[\mathbf{1}_{\{Y^\mu(t+h)=j, t_1 > t\}} | Y^\mu(s), s \leq t]] dt \\ &= E_\mu \int_0^\infty E[\mathbf{1}_{\{Y^\mu(t+h)=j, t_1 > t\}} | Y^\mu(s), s \leq t] dt \\ &= E_\mu \int_0^\infty \mathbf{1}_{\{t_1 > t\}} \cdot P_{Y^\mu(t)j}^\mu(h) dt \\ &= \sum_{k \in \mathbb{N}} P_{kj}^\mu(h) \cdot E_\mu \int_0^{t_1} \mathbf{1}_{\{Y^\mu(t)=k\}} dt \end{aligned}$$

and this shows that ν is indeed an invariant measure for Y^μ . Because $\sum_{k \in \mathbb{N}} \nu(k) = \int_0^\infty \sum_{k \in \mathbb{N}} P_\mu(X(t)=k) dt = \int_0^\infty P_\mu(R > t) dt = E_\mu R < \infty$, we can normalize ν to a probability distribution. So $\Phi\mu$ is the unique invariant distribution of $Y^\mu(t)$. \square

4.3. Quasi-stationary distribution as fixpoint

We summarize the results of Pakes [18], in order to show that under our basic assumptions a fixpoint of Φ is a quasi-stationary distribution.

Pakes has the following setting, that is contained in ours:

There is a countable state space $S = N_0$ and a stable and conservative q -matrix Q on S . 0 is the only absorbing state and $C = \mathbb{N}$ is irreducible for the minimal Q -function and hence for any Q -function. It is further assumed, that there is an $i \in C$ for which $q(i, 0) > 0$.

Definition 4.14. Let P be some Q -function and $(\mu(j))_{j \in C}$ be a probability distribution on C and define $P(\cdot) = (P_j(\cdot))_{j \in S}$ by

$$P_j(t) = \sum_{i \in C} \mu(i) P_{ij}(t), j \in S, t \geq 0.$$

Then μ is a **quasi-stationary distribution on C for P** if for all $t > 0$,

$$\frac{P_j(t)}{\sum_{i \in C} P_i(t)} = \mu(j), j \in C.$$

4. The Main Theorem

Definition 4.15. Let $\beta \geq 0$ be some non-negative real number. A collection of positive numbers $\mu = (\mu(j), j \in C)$ is called a **β -subinvariant measure on C for Q** if

$$\sum_{i \in C} \mu(i) q(i, j) \leq -\beta \mu(j), j \in C, \quad (4.13)$$

and **β -invariant** if equality holds for all $j \in C$.

Definition 4.16. μ is said to be **β -subinvariant on C for P** , where P is a transition function, if

$$\sum_{i \in C} \mu(i) P_{ij}(t) \leq e^{-\beta t} \mu(j), j \in C, t \geq 0, \quad (4.14)$$

and **β -invariant** if equality holds for all $j \in C$ and $t \geq 0$.

A given measure μ is **convergent** if $\sum_{i \in C} \mu(i) < \infty$.

Proposition 4.17. Let P be an arbitrary Q -function and suppose that μ is a proper probability distribution on C . Then μ is a quasi-stationary distribution on C for P if and only if, for some $\beta > 0$, μ is β -invariant on C for P .

Proof. Firstly, suppose that μ is β -invariant for some $\beta > 0$. Then with the notation as in the definition 4.14, we have for all $j \in C$ and $t > 0$

$$\frac{P_j(t)}{\sum_{i \in C} P_i(t)} = \frac{e^{-\beta t} \mu(j)}{\sum_{i \in C} e^{-\beta t} \mu(i)} = \frac{\mu(j)}{\sum_{i \in C} \mu(i)} = \mu(j),$$

and so μ is a quasi-stationary distribution on C for P .

On the other hand, suppose that μ is a quasi-stationary distribution on C for P . With the definition $g(t) = \sum_{i \in C} P_i(t)$, we have because μ is a quasi-stationary distribution

$$P_j(t) = \sum_{i \in C} \mu(i) P_{ij}(t) = g(t) \mu(j), j \in C, t > 0. \quad (4.15)$$

So it suffices to prove, that there is some $\beta > 0$ with $g(t) = e^{-\beta t}$. We therefore check, that g satisfies the Cauchy exponential functional equation $g(t+s) = g(t)g(s)$. By Chapman-Kolmogorov and bearing in mind, that 0 is an absorbing state, we get

$$\begin{aligned} P_j(s+t) &= \sum_{i \in C} \mu(i) P_{ij}(s+t) = \sum_{i \in C} \mu(i) \sum_{k \in C} P_{ik}(s) P_{kj}(t) \\ &= \sum_{k \in C} P_{kj}(t) \sum_{i \in C} \mu(i) P_{ik}(s) = \sum_{k \in C} P_{kj}(t) P_k(s). \end{aligned}$$

Combining this with (4.15), gives

$$\begin{aligned} g(t+s) \mu(j) &= P_j(s+t) = \sum_{k \in C} P_{kj}(t) P_k(s) \\ &= \sum_{k \in C} P_{kj}(t) g(s) \mu(k) = g(s) P_j(t) \end{aligned}$$

4. The Main Theorem

and summing over $j \in C$ gives $g(t+s) = g(s)g(t)$. We must have $0 < g(t) \leq 1$, because for some arbitrary $j \in C$ we have $g(t) \geq P_j(t) \geq \mu(j)P_{jj}(t) > 0$ and

$$g(t) = \sum_{j \in C} \sum_{i \in C} \mu(i) P_{ij}(t) = \sum_{i \in C} \mu(i) \sum_{j \in C} P_{ij}(t) \leq 1.$$

Since g is continuous, it follows that there is some $\beta \geq 0$ such that $g(t) = e^{-\beta t}$. Because we assumed, there exist a least one $i \in C$ with $P_{i0}(t) > 0$ for all $t > 0$ it's $g(t) < 1$ and so we refuted the case $\beta = 0$. □

The following theorem is a simple extension of Pollet and Vere-Jones (1992), so the proof will be omitted.

Theorem 4.18. *If μ is a β -subinvariant measure on C for P then μ is a β -subinvariant measure on C for Q . A necessary condition for μ to be a μ -invariant measure on C for Q is that P satisfies the Kolmogorov forward equations over C . If μ is β -invariant on C for P , then this condition is also sufficient.*

The last two theorems give the following simple result.

Corollary 4.19. *Suppose that μ is a quasi-stationary distribution on C for P . Then for some $\beta > 0$, the distribution μ is a β -subinvariant measure on C for Q and β -invariant if and only if P satisfies the Kolmogorov forward equations over C .*

Proof. Proposition 4.17 tells there is some $\beta > 0$ such that μ is a β -invariant measure on C for P . Then the last theorem tells that μ is also a β -subinvariant measure on C for Q and β -invariant on C for Q iff P satisfies the Kolmogorov forward equations over C . □

We define for $i \in C$

$$a_i^P = \lim_{t \rightarrow \infty} P_{i0}(t).$$

Of course, this limit does exit, because $P_{i0}(t)$ is increasing in t because 0 is an absorbing state.

Lemma 4.20. *Let P be a Q -function and suppose that, for some $\beta > 0$, there exists a convergent β -subinvariant measure on C for P . Then $a_i^P = h_i^P$ for all $i \in C$ where*

$$h_i^P = \lim_{t \rightarrow \infty} \sum_{j \in S} P_{ij}(t), i \in C.$$

Proof. So let μ be a convergent β -subinvariant measure on C for P with $\beta > 0$. We have in particular for all $i, j \in C$ that $\mu(i)P_{ij}(t) \leq e^{-\beta t}\mu(j)$. If we recall again that per definition all $\mu(i) > 0$, then we get for all $i \in C$

$$\sum_{j \in C} P_{ij}(t) \leq e^{-\beta t} \frac{1}{\mu(i)} \sum_{j \in C} \mu(j),$$

4. The Main Theorem

and hence because μ is convergent and $\beta > 0$

$$\lim_{t \rightarrow \infty} \sum_{j \in C} P_{ij}(t) = 0.$$

The result follows. \square

Theorem 4.21. *Let P be a Q -function and suppose that, for some $\beta > 0$, μ is a convergent β -invariant measure on C for P . Then*

$$\beta \geq \frac{\sum_{i \in C} \mu(i) q(i, 0)}{\sum_{i \in C} \mu(i) a_i^P}.$$

Proof. By the forward inequalities, and recalling that 0 is an absorbing state (which means that $q(0, 0) = 0$) we get

$$P'_{i0}(t) \geq \sum_{k \in S} P_{ik}(t) q(k, 0) \geq \sum_{k \in C} P_{ik}(t) q(k, 0), i \in C.$$

Integrating both sides yields

$$P_{i0}(t) \geq \sum_{k \in C} \int_0^t P_{ik}(s) q(k, 0) ds, i \in C.$$

On multiplying by $\mu(i)$ and summing over $i \in C$, we get

$$\sum_{i \in C} \mu(i) P_{i0}(t) \geq \sum_{i \in C} \mu(i) \sum_{k \in C} \int_0^t P_{ik}(s) q(k, 0) ds.$$

Then using that μ is convergent and a β -invariant measure on C for P we follow by

$$\begin{aligned} \infty &> \sum_{i \in C} \mu(i) \geq \sum_{i \in C} \mu(i) P_{i0}(t) \\ &\geq \sum_{i \in C} \mu(i) \sum_{k \in C} \int_0^t P_{ik}(s) q(k, 0) ds \\ &= \sum_{k \in C} q(k, 0) \int_0^t \sum_{i \in C} \mu(i) P_{ik}(s) ds \\ &= \sum_{k \in C} q(k, 0) \int_0^t e^{-\beta s} \mu(k) ds \\ &= \frac{1}{\beta} (1 - e^{-\beta t}) \sum_{k \in C} \mu(k) q(k, 0), \end{aligned} \tag{4.16}$$

that $\sum_{k \in C} \mu(k) q(k, 0) < \infty$. Now we introduce the *dishonesty function*

$$e_i^P(t) = 1 - \sum_{j \in S} P_{ij}(t), i \in S, t \geq 0,$$

4. The Main Theorem

and note that the limit of the dishonesty function $e_i^P = \lim_{t \rightarrow \infty} e_i^P(t)$ exists because of the previous lemma and $e_i^P = 1 - h_i^P$. Then we get

$$\begin{aligned}
 (1 - e^{-\beta t}) \sum_{i \in C} \mu(i) - \sum_{i \in C} \mu(i) e_i^P(t) &= (1 - e^{-\beta t}) \sum_{i \in C} \mu(i) - \sum_{i \in C} \mu(i) + \sum_{i \in C} \mu(i) \sum_{j \in S} P_{ij}(t) \\
 &= -e^{-\beta t} \sum_{i \in C} \mu(i) + \sum_{j \in C} \underbrace{\sum_{i \in C} \mu(i) P_{ij}(t)}_{=e^{-\beta t} \mu(j)} + \sum_{i \in C} \mu(i) P_{i0}(t) \\
 &= \sum_{i \in C} \mu(i) P_{i0}(t).
 \end{aligned} \tag{4.17}$$

Combining the results so far, we get

$$\begin{aligned}
 \sum_{i \in C} \mu(i) e_i^P(t) &\stackrel{(4.17)}{=} (1 - e^{-\beta t}) \sum_{i \in C} \mu(i) - \sum_{i \in C} \mu(i) P_{i0}(t) \\
 &\stackrel{(4.16)}{\leq} (1 - e^{-\beta t}) \sum_{i \in C} \mu(i) - \frac{1}{\beta} (1 - e^{-\beta t}) \sum_{i \in C} \mu(i) q(i, 0) \\
 &= \frac{1}{\beta} (1 - e^{-\beta t}) \left(\beta \sum_{i \in C} \mu(i) - \sum_{i \in C} \mu(i) q(i, 0) \right).
 \end{aligned}$$

If we let $t \rightarrow \infty$ and using bounded convergence, we get

$$\beta \sum_{i \in C} \mu(i) \geq \sum_{i \in C} \mu(i) q(i, 0) + \beta \sum_{i \in C} \mu(i) e_i^P.$$

But by definition, $e_i^P = 1 - h_i^P$ and by the previous Lemma $h_i^P = a_i^P$ for all $i \in C$. Thus follows

$$\beta \sum_{i \in C} \mu(i) a_i^P \geq \sum_{i \in C} \mu(i) q(i, 0).$$

Because we assumed there is some $i \in C$ with $q(i, 0) > 0$, the right-hand side of the above inequality is positive and the result follows. \square

Theorem 4.22. *Let P be a Q -function which satisfies the Kolmogorov forward equations over C . Suppose further that for some $\beta > 0$, μ is a convergent β -subinvariant measure on C for P . Then*

$$\beta \sum_{i \in C} \mu(i) a_i^P \leq \sum_{i \in C} \mu(i) q(i, 0),$$

and if P is honest

$$\beta \leq \frac{\sum_{i \in C} \mu(i) q(i, 0)}{\sum_{i \in C} \mu(i)}.$$

4. The Main Theorem

Proof. The difference of the assumptions to the previous theorem is that here, the Kolmogorov forward equations (and not only the inequalities) are satisfied and that on the other hand, μ is only a *subinvariant* measure on C for P (and not invariant). The result then follows by repeating the same calculations and arguments as in the previous theorem. \square

The two previous theorems combined give the following corollary.

Corollary 4.23. *Let P be a Q -function which satisfies the Kolmogorov forward equations over C and suppose that for some $\beta > 0$, μ is a convergent β -invariant measure on C for P . Then*

$$\beta = \frac{\sum_{i \in C} \mu(i) q(i, 0)}{\sum_{i \in C} \mu(i) a_i^P}.$$

Moreover,

$$\beta = \frac{\sum_{i \in C} \mu(i) q(i, 0)}{\sum_{i \in C} \mu(i)}$$

if and only if P is honest.

Theorem 4.24. *Suppose that μ is a quasi-stationary distribution on C for P and that P satisfies the Kolmogorov forward equations over S . Then μ is a β -invariant measure on C for Q , where*

$$\beta = \frac{\sum_{i \in C} \mu(i) q(i, 0)}{\sum_{i \in C} \mu(i) a_i^P},$$

and

$$\beta = \sum_{i \in C} \mu(i) q(i, 0)$$

if and only if P is honest.

Proof. Since μ is a quasi-stationary distribution on C for P , after Proposition 4.17 μ is also a convergent β -invariant measure on C for P for some $\beta > 0$. After the last Corollary, because P satisfies the Kolmogorov forward equations $\beta = \frac{\sum_{i \in C} \mu(i) q(i, 0)}{\sum_{i \in C} \mu(i) a_i^P}$ and $\beta = \sum_{i \in C} \mu(i) q(i, 0)$ if and only if P is honest. But then after Theorem 4.18 since P satisfies the Kolmogorov forward equations μ is also a β -invariant measure on C for Q . \square

Theorem 4.25. *Let $F_{ij}(t)$ be the minimal Q -function and suppose that, for some $\beta > 0$, μ is a convergent β -subinvariant measure on C for Q . Then μ is a β -invariant measure on C for F if and only if*

$$\beta \sum_{i \in C} \mu(i) a_i^F = \sum_{i \in C} \mu(i) q(i, 0). \quad (4.18)$$

Proof. See [16] Theorem 4.1. on page 91. \square

Theorem 4.26. *Let μ be a probability distribution on \mathbb{N} with $E_\mu R < \infty$. If μ is a fixpoint of Φ , then μ is a quasi-stationary distribution for $X(t)$.*

4. The Main Theorem

Proof. We know from the previous subsection that $\mu(y) > 0$ for all $y \in \mathbb{N}$ and that μ satisfies the equation $\mu Q^\mu = 0$. Therefore for all $j \in C$, we have

$$\sum_{i \in C} \mu(i) q^\mu(i, j) = \sum_{i \in C} \mu(i) (q(i, j) + q(i, 0) \mu(j)) = 0.$$

So for all $j \in C$ holds

$$\sum_{i \in C} \mu(i) q(i, j) = - \left(\sum_{i \in C} \mu(i) q(i, 0) \right) \mu(j).$$

This means that μ is a β -invariant measure on C for Q with $\beta = \sum_{i \in C} \mu(i) q(i, 0)$. We see $\beta > 0$, because $\mu(i) > 0$ for all $i \in \mathbb{N}$ and since absorption is certain, there must be some i such that $q(i, 0) > 0$.

We recall $X(t)$ equals the minimal process, so its transition function F is the minimal Q -function. In addition, we have for all $i \in C$, since absorption is certain

$$a_i^F = \lim_{t \rightarrow \infty} F_{i0}(t) = \lim_{t \rightarrow \infty} P_i(X(t) = 0) = \lim_{t \rightarrow \infty} P_i(R \leq t) = P_i(R < \infty) = 1.$$

Also $\beta \sum_{i \in C} \mu(i) a_i^F = \beta \sum_{i \in C} \mu(i) \cdot 1 = \beta = \sum_{i \in C} \mu(i) q(i, 0)$.

Then because of Theorem 4.25, μ is also a β -invariant measure on C for F .

Then Proposition 4.17 shows, that μ is indeed a quasi-stationary distribution. \square

4.4. Checking the assumptions of Schauder-Tychonov fixpoint theorem

We already outlined, how we want to apply the Schauder-Tychonov fixpoint theorem to prove the existence of a fixpoint of Φ , which is a quasi-stationary distribution. In this section, we assume there exists some $\theta > 0$ such that \mathcal{M}_θ as defined in (4.9) is not empty, i.e.

$$K = \mathcal{M}_\theta = \{\mu : \mu \text{ probability measure on } \mathbb{N} \text{ such that } P_\mu(R > t) = e^{-\frac{t}{\theta}}\} \neq \emptyset,$$

and check the list of assumptions from the Schauder-Tychonov fixpoint theorem.

Theorem 4.27. *Let (4.2) hold, i.e.*

$$\lim_{x \rightarrow \infty} P_x(R < t) = 0, \quad \forall t \geq 0.$$

Then if $\mathcal{M}_\theta \neq \emptyset$, the transformation Φ has a fixpoint in \mathcal{M}_θ .

Proof. **1. $\Phi : \mathcal{M}_\theta \rightarrow \mathcal{M}_\theta$ is well defined.**

4. The Main Theorem

We want to show $\Phi(\mathcal{M}_\theta) \subseteq \mathcal{M}_\theta$. If $\mu \in \mathcal{M}_\theta$, then $E_\mu R = \theta < \infty$ (as the expectation of the exponential distribution with parameter $1/\theta$).

$$\begin{aligned}
P_{\Phi\mu}(R > s) &= \sum_{y \in \mathbb{N}} P_y(R > s) \Phi\mu(y) \\
&= \sum_{y \in \mathbb{N}} P_y(R > s) \frac{1}{E_\mu[R]} \int_0^\infty P_\mu(X(t) = y) dt \\
&= \frac{1}{E_\mu[R]} \int_0^\infty \sum_{y \in \mathbb{N}} P_\mu(X(t) = y) P(R > s + t | X(t) = y) dt \\
&= \frac{1}{E_\mu[R]} \int_0^\infty P_\mu(R > t + s) dt \\
&= \frac{1}{E_\mu R} \int_s^\infty P_\mu(R > t) dt \\
&= \frac{1}{\theta} \int_s^\infty e^{-\frac{t}{\theta}} dt \\
&= \int_{\frac{s}{\theta}}^\infty e^{-t} dt = e^{-\frac{s}{\theta}}.
\end{aligned}$$

So $\Phi\mu \in \mathcal{M}_\theta$.

2. \mathcal{M}_θ is convex.

So let $\mu, v \in \mathcal{M}_\theta$ and $\lambda \in [0, 1]$, then

$$\begin{aligned}
P_{\lambda\mu + (1-\lambda)v}(R > t) &= \sum_{y \in \mathbb{N}} [\lambda\mu(y) + (1-\lambda)v(y)] P_y(R > t) \\
&= \lambda \sum_{y \in \mathbb{N}} \mu(y) P_y(R > t) + (1-\lambda) \sum_{y \in \mathbb{N}} v(y) P_y(R > t) \\
&= \lambda P_\mu(R > t) + (1-\lambda) P_v(R > t) \\
&= \lambda e^{-\frac{t}{\theta}} + (1-\lambda) e^{-\frac{t}{\theta}} = e^{-\frac{t}{\theta}}.
\end{aligned}$$

So \mathcal{M}_θ is convex.

3. \mathcal{M}_θ is compact.

For compactness, we show sequential compactness. So let $(\mu_k)_{k \in \mathbb{N}} \subseteq \mathcal{M}_\theta$. For every $\epsilon > 0$ there exists $t_\epsilon > 0$ such that $e^{-\frac{t_\epsilon}{\theta}} < \epsilon$. So for all $k \in \mathbb{N}$ holds

4. The Main Theorem

$$\begin{aligned} \epsilon &\geq P_{\mu_k}(R > t_\epsilon) = \sum_{y \in \mathbb{N}} \mu_k(y) P_y(R > t_\epsilon) \\ &\geq \sum_{\{y \in \mathbb{N} : P_y(R > t_\epsilon) > \frac{1}{2}\}} \mu_k(y) P_y(R > t_\epsilon) \geq \frac{1}{2} \mu_k\{y \in \mathbb{N} : P_y(R > t_\epsilon) > \frac{1}{2}\} \end{aligned}$$

and consequently for all $k \in \mathbb{N}$, $2\epsilon \geq \mu_k\{y \in \mathbb{N} : P_y(R \leq t_\epsilon) < \frac{1}{2}\}$. Because $\lim_{y \rightarrow \infty} P_y(R \leq t_\epsilon) = 0$, there exists $Y_\epsilon \in \mathbb{N}$ such that for all $y \geq Y_\epsilon$, $P_y(R \leq t_\epsilon) < \frac{1}{2}$. So for all $k \in \mathbb{N}$, $\mu_k\{y \geq Y_\epsilon\} \leq 2\epsilon$ and this proves the tightness of the family $(\mu_k)_{k \in \mathbb{N}}$.

So after the Theorem of Prohorov (see for instance Klenke [14] Theorem 13.29 on page 261), there is a subsequence $(k_j)_{j \in \mathbb{N}}$ such that

$$\mu_{k_j} \Rightarrow \mu \text{ as } j \rightarrow \infty.$$

for some probability distribution μ on \mathbb{N} . Then

$$\begin{aligned} P_\mu(R > t) &= \sum_{y \in \mathbb{N}} \mu(y) P_y(R > t) \\ &= \lim_{Y \rightarrow \infty} \lim_{j \rightarrow \infty} \sum_{y=1}^Y \mu_{k_j}(y) P_y(R > t) \end{aligned}$$

Now we would like to exchange the limits and apply therefore Theorem A.3. The limit $\lim_{Y \rightarrow \infty} \sum_{y=1}^Y \mu_{k_j}(y) P_y(R > t)$ is reached uniformly in j . Indeed, let $\epsilon > 0$ be arbitrary. Since the family $(\mu_k)_{k \in \mathbb{N}}$ is tight, there exists Y_ϵ such that $\mu_{k_j}\{y \geq Y_\epsilon\} \leq \epsilon$ for all $j \in \mathbb{N}$. Thus we have for all $Y \geq Y_\epsilon$, $j \in \mathbb{N}$,

$$\begin{aligned} &\left| \sum_{y \in \mathbb{N}} \mu_{k_j}(y) P_y(R > t) - \sum_{y=1}^Y \mu_{k_j}(y) P_y(R > t) \right| \\ &= \sum_{y=Y+1}^{\infty} \mu_{k_j}(y) P_y(R > t) \\ &\leq \mu_{k_j}\{y \geq Y_\epsilon\} \leq \epsilon. \end{aligned} \tag{4.19}$$

Thus

$$\begin{aligned} P_\mu(R > t) &= \lim_{j \rightarrow \infty} \lim_{Y \rightarrow \infty} \sum_{y=1}^Y \mu_{k_j}(y) P_y(R > t) \\ &= \lim_{j \rightarrow \infty} P_{\mu_{k_j}}(R > t) = e^{-\frac{t}{\theta}}. \end{aligned}$$

and this proves that $\mu \in \mathcal{M}_\theta$ and thus the sequential compactness of \mathcal{M}_θ .

4. The Main Theorem

4. $\Phi : \mathcal{M}_\theta \rightarrow \mathcal{M}_\theta$ is continuous.

If $\mu_k(z) \rightarrow \mu(z)$ for all $z \geq 1$ and $\mu_k, \mu \in \mathcal{M}_\theta$, then $E_{\mu_k} R = E_\mu R = \theta$. Then for any $A \subseteq \mathbb{N}$, by Fatous lemma

$$\begin{aligned}
 \liminf_{k \rightarrow \infty} \sum_{y \in A} \Phi \mu_k(y) &= \liminf_{k \rightarrow \infty} \sum_{y \in A} \frac{1}{E_{\mu_k} R} \int_0^\infty P_{\mu_k}(X(t) = y) dt \\
 &= \liminf_{k \rightarrow \infty} \sum_{y \in A} \frac{1}{\theta} \int_0^\infty \sum_{z \in \mathbb{N}} \mu_k(z) P_z(X(t) = y) dt \\
 &\geq \sum_{y \in A} \frac{1}{\theta} \int_0^\infty \sum_{z \in \mathbb{N}} \liminf_{k \rightarrow \infty} \mu_k(z) P_z(X(t) = y) dt \quad (4.20) \\
 &= \sum_{y \in A} \frac{1}{\theta} \int_0^\infty \sum_{z \in \mathbb{N}} \mu(z) P_z(X(t) = y) dt \\
 &= \sum_{y \in A} \Phi \mu(y).
 \end{aligned}$$

But since $1 = \liminf_{k \rightarrow \infty} \sum_{y \in \mathbb{N}} \Phi \mu_k(y) \geq \liminf_{k \rightarrow \infty} \sum_{y \in A} \Phi \mu_k(y) + \liminf_{k \rightarrow \infty} \sum_{y \in A^c} \Phi \mu_k(y) \geq \sum_{y \in A} \Phi \mu(y) + \sum_{y \in A^c} \Phi \mu(y) = 1$, we must have equality in (4.20). Then also

$$\begin{aligned}
 1 &= \limsup_{k \rightarrow \infty} \sum_{z \in \mathbb{N}} \Phi \mu_k(z) = \limsup_{k \rightarrow \infty} \{ \Phi \mu_K(y) + \sum_{z \neq y} \Phi \mu_K(z) : K \geq k \} \\
 &\geq \limsup_{k \rightarrow \infty} \{ \Phi \mu_K(y) : K \geq k \} + \inf_{z \neq y} \{ \sum_{z \neq y} \Phi \mu_K(z) : K \geq k \} \\
 &= \limsup_{k \rightarrow \infty} \Phi \mu_k(y) + \liminf_{k \rightarrow \infty} \sum_{z \neq y} \Phi \mu_k(z) = \limsup_{k \rightarrow \infty} \Phi \mu_k(y) + \sum_{z \neq y} \Phi \mu(z)
 \end{aligned}$$

so in total

$$\Phi \mu(y) = \liminf_{k \rightarrow \infty} \Phi \mu_k(y) \leq \limsup_{k \rightarrow \infty} \Phi \mu_k(y) \leq \Phi \mu(y)$$

which gives the desired continuity. □

4.5. Associated renewal processes

To complete the proof of Theorem 4.6, we only need to show that there is some $\theta > 0$, such that $\mathcal{M}_\theta \neq \emptyset$. That is what is left to this subsection and in particular to Theorem 4.30. It will be amazing to see, how we can apply our results from renewal theory here.

Recall that F^μ is the distribution of the absorption time R under the initial distribution μ . Recall also the definition of the map Ψ in (3.2) as the limit distribution of the remaining life time.

4. The Main Theorem

Lemma 4.28. *If $E_\mu[R] < \infty$, then $F^{\Phi\mu} = \Psi F^\mu$.*

Proof.

$$\begin{aligned}
1 - F^{\Phi\mu}(s) &= P_{\Phi\mu}(R > s) \\
&= \sum_{y \in \mathbb{N}} P_y(R > s) \Phi\mu(y) \\
&= \sum_{y \in \mathbb{N}} P_y(R > s) \frac{1}{E_\mu[R]} \int_0^\infty P_\mu(X(t) = y) dt \\
&= \frac{1}{E_\mu[R]} \int_0^\infty \sum_{y \in \mathbb{N}} P_\mu(X(t) = y) P(R > s + t | X(t) = y) dt \\
&= \frac{1}{E_\mu[R]} \int_0^\infty P_\mu(R > t + s) dt \\
&= \frac{1}{E_\mu R} \int_s^\infty P_\mu(R > t) dt \\
&= \frac{1}{E_\mu[R]} \int_s^\infty 1 - F^\mu(t) dt \\
&= 1 - \Psi F^\mu(s).
\end{aligned} \tag{4.21}$$

□

Corollary 4.29. *If $E_\mu[R] < \infty$, then for all $n \in \mathbb{N}$, $F^{\Phi^n\mu} = \Psi^n F^\mu$.*

Proof. This follows by the previous lemma and induction over n . □

Theorem 4.30. *Assume all assumptions from main theorem, i.e. assume (4.2) and (4.3) hold. Recall the definition of the exponential rate of survival*

$$\lambda_0 = \sup\{\lambda : E_x[e^{\lambda R}] < \infty\}. \tag{4.22}$$

Then λ_0 is independent of the choice of x and for each x

$$F_n(t) := F^{\Phi^n\delta_x}(t) = P_{\Phi^n\delta_x}(R \leq t) \rightarrow 1 - e^{-t\lambda_0}$$

with

$$0 < \lambda_0 = \left(\lim_{n \rightarrow \infty} E_{\Phi^n\delta_x} R \right)^{-1} < \infty. \tag{4.23}$$

Finally, for fixed $x \geq 1$, there exists a subsequence $(n_j)_{j \in \mathbb{N}}$ and a probability measure μ_∞ on \mathbb{N} such that $\Phi^{n_j}\delta_x$ converges weakly to μ_∞ and such that

$$P_{\mu_\infty}(R > t) = e^{-t\lambda_0}. \tag{4.24}$$

Remark 4.31. *The theorem shows, that for $\theta = \frac{1}{\lambda_0} > 0$, we have $\mu_\infty \in \mathcal{M}_\theta \neq \emptyset$.*

4. The Main Theorem

Proof. The independence of x follows from irreducibility of Q , see also (4.7).

Fix now $x \in \mathbb{N}$ and write F for $F_0 = F^{\delta_x}$, so $F(t) = P_x(R \leq t)$. We note that F is also an aperiodic interarrival distribution. Furthermore, all moments of F exist because for any $k \in \mathbb{N}$ and $r > 0$ large enough, we have $r^k < e^{\lambda r}$. So $E_x[R^k] = E_x[R^k \cdot \mathbf{1}_{\{R < r\}}] + E_x[R^k \cdot \mathbf{1}_{\{R \geq r\}}] \leq r^k P_x(R < r) + E_x[e^{\lambda R} \cdot \mathbf{1}_{\{R \geq r\}}] \leq r^k + E_x[e^{\lambda R}] < \infty$. Because of the last corollary $F_n = F^{\Phi^n \delta_x} = \Psi^n F^{\delta_x} = \Psi^n F$ and so our results in the chapter about renewal theory are applicable to the F_n .

At first, we prove

$$\liminf_{n \rightarrow \infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} \geq \frac{1}{\lambda_0}. \quad (4.25)$$

If this would fail, then by Lemma 3.6, one can find a subsequence $(n_j)_{j \in \mathbb{N}}$ such that F_{n_j} converges to an exponential distribution with mean $\theta < \frac{1}{\lambda_0}$.

But we show now, that this contradicts the maximality of λ_0 . Set $\mu_n = \Phi^n \delta_x$. Then the family

$$\{\mu_{n_j} : j \in \mathbb{N}\} \quad (4.26)$$

is tight. To see this, let $\epsilon > 0$ be arbitrary. There is t_ϵ^* such that $\lim_{j \rightarrow \infty} P_{\mu_{n_j}}(R > t_\epsilon^*) = e^{-t_\epsilon^* \frac{1}{\theta}} < \epsilon$. Then there is $J \in \mathbb{N}$ such that for all $j > J$, $P_{\mu_{n_j}}(R > t_\epsilon^*) < \epsilon$. Because absorption is sure and because of continuity from below of a probability measure, we have for $j \in \{1, \dots, J\}$, $\lim_{t \rightarrow \infty} P_{\mu_{n_j}}(R < t) = P_{\mu_{n_j}}(R < \infty) = 1$. So there is $t_{j,\epsilon}$ such that $P_{\mu_{n_j}}(R > t_{j,\epsilon}) < \epsilon$. Then with $t_\epsilon = \max\{t_\epsilon^*, t_{1,\epsilon}, \dots, t_{J,\epsilon}\}$, we have for all $j \in \mathbb{N}$,

$$\begin{aligned} \epsilon &\geq P_{\mu_{n_j}}(R > t_\epsilon) = \sum_{y \in \mathbb{N}} \mu_{n_j}(y) P_y(R > t_\epsilon) \\ &\geq \sum_{\{y \in \mathbb{N} : P_y(R > t_\epsilon) > \frac{1}{2}\}} \mu_{n_j}(y) P_y(R > t_\epsilon) \geq \frac{1}{2} \mu_{n_j} \{y \in \mathbb{N} : P_y(R > t_\epsilon) > \frac{1}{2}\} \end{aligned}$$

and consequently for all $j \in \mathbb{N}$, $2\epsilon \geq \mu_{n_j} \{y \in \mathbb{N} : P_y(R \leq t_\epsilon) < \frac{1}{2}\}$. Because $\lim_{y \rightarrow \infty} P_y(R \leq t_\epsilon) = 0$, there exists $Y_\epsilon \in \mathbb{N}$ such that for all $y \geq Y_\epsilon$, $P_y(R \leq t_\epsilon) < \frac{1}{2}$. So for all $j \in \mathbb{N}$, $\mu_{n_j} \{y \geq Y_\epsilon\} \leq 2\epsilon$ and this proves the tightness.

So after the Theorem of Prohorov, by going to a sub-subsequence if necessary, we may assume

$$\mu_{n_j} \Rightarrow \mu_\infty \text{ as } j \rightarrow \infty,$$

for some probability distribution μ_∞ on \mathbb{N} . Then because of Theorem A.3 and the tightness, just as in (4.19), we are able to exchange limits and get

4. The Main Theorem

$$\begin{aligned} P_{\mu_\infty}(R > t) &= \sum_{y \in \mathbb{N}} \mu_\infty(y) P_y(R > t) = \lim_{j \rightarrow \infty} \sum_{y \in \mathbb{N}} \mu_{n_j}(y) P_y(R > t) \\ &= \lim_{j \rightarrow \infty} (1 - F_{n_j}(t)) = e^{-\frac{t}{\theta}}. \end{aligned}$$

μ_∞ charges some point y and $\mu_\infty(y) P_y(R > t) \leq P_{\mu_\infty}(R > t)$, thus

$$P_y(R > t) \leq \frac{1}{\mu_\infty(y)} e^{-\frac{t}{\theta}}$$

which contradicts the maximality of λ_0 , because let $\kappa > 0$ with $\lambda_0 < \frac{1}{\kappa} < \frac{1}{\theta}$ then a contradiction would be if $E_y \left[e^{\frac{1}{\kappa} R} \right] < \infty$ and this is so because $E_y \left[e^{\frac{1}{\kappa} R} \right] = 1 + E_y \left[e^{\frac{1}{\kappa} R} - 1 \right] = 1 + \int_0^\infty P_y \left(e^{\frac{1}{\kappa} R} - 1 > t \right) dt = 1 + \int_0^\infty P_y(R > \kappa \ln(t+1)) dt \leq 1 + \frac{1}{\mu_\infty(y)} \int_0^\infty e^{-\frac{\kappa \ln(t+1)}{\theta}} dt = 1 + \frac{1}{\mu_\infty(y)} \int_1^\infty t^{-\frac{\kappa}{\theta}} dt < \infty$. So (4.25) must hold.

Now we prove, that there exists μ_∞ such that (4.24) holds, which works similarly to what we just did. Because of Remark 4.9, for all $\epsilon > 0$, there exist a constant $C_\epsilon > 0$ such that

$$P_x(R > t) \leq C_\epsilon e^{-t(\lambda_0 - \epsilon)}.$$

Then because of Theorem 3.7, we can find a subsequence $(n_j)_{j \in \mathbb{N}}$ such that F_{n_j} converges to an exponential distribution with mean $\theta \leq \frac{1}{\lambda_0 - \epsilon}$. By doing this for $\epsilon \rightarrow 0$, we can get a subsequence $(n_j)_{j \in \mathbb{N}}$ such that F_{n_j} converges to an exponential distribution with mean $\theta \leq \frac{1}{\lambda_0}$. If $\theta < \frac{1}{\lambda_0}$, then we get a contradiction to the maximality of λ_0 just as before. So $\theta = \frac{1}{\lambda_0}$ and the argumentation before also provides us with μ_∞ such that $P_{\mu_\infty}(R > t) = e^{-t\lambda_0}$ for all $t \geq 0$. We also note, that from the existence of μ_∞ follows

$$P_x(R > t) \leq C e^{-\lambda_0 t}.$$

Indeed, there exists $y \in \mathbb{N}$ in the support of μ_∞ , so we get $P_y(R > t) \leq \frac{1}{\mu_\infty(y)} \cdot e^{-\lambda_0 t}$ for all $t \geq 0$. Because \mathbb{N} is an irreducible class, there is $s > 0$ such that $P_y(X(s) = x) > 0$. Then we have for all $t \geq s$,

$$\begin{aligned} \frac{1}{\mu_\infty(y)} \cdot e^{-\lambda_0 t} &\geq P_y(R > t) \geq P_y(R > t, X(s) = x) \\ &= P_y(X(s) = x) \cdot P(R > t | X(s) = x) = P_y(X(s) = x) \cdot P_x(R > t - s). \end{aligned}$$

So we have for all $t \geq 0$, $P_x(R > t) \leq \frac{e^{-\lambda_0 s}}{\mu_\infty(y) P_y(X(s) = x)} \cdot e^{-\lambda_0 t}$.

Next we prove a converse of (4.25), namely

$$\limsup_{n \rightarrow \infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} \leq \frac{1}{\lambda_0}. \quad (4.27)$$

4. The Main Theorem

Suppose this is not true. Then there exists $\eta > 0$ and a subsequence $(n_l)_{l \in \mathbb{N}}$ such that

$$\frac{m_{n_l+1}(F)}{(n_l+1)m_{n_l}(F)} \geq \frac{1}{\lambda_0} + \eta \quad (4.28)$$

for all $l \in \mathbb{N}$. For C satisfying

$$\frac{1}{1+C\eta} \left(\frac{1}{\lambda_0} + \eta \right) \geq \frac{1}{\lambda_0} + \frac{\eta}{2}$$

we have for $n_l \leq j < j+1 \leq n_l(1+C\eta)$,

$$\begin{aligned} \frac{m_{j+1}(F)}{(j+1)m_j(F)} &\geq \frac{m_{n_l+1}(F)}{(j+1)m_{n_l}(F)} = \frac{m_{n_l+1}(F)}{(n_l+1)m_{n_l}(F)} \frac{n_l+1}{j+1} \\ &\geq \frac{m_{n_l+1}(F)}{(n_l+1)m_{n_l}(F)} \frac{n_l+1}{n_l(1+C\eta)} \geq \left(\frac{1}{\lambda_0} + \eta \right) \frac{1}{1+C\eta} \geq \frac{1}{\lambda_0} + \frac{\eta}{2}. \end{aligned}$$

For shorter notation, set $i(n_l) = \lfloor n_l(1+C\eta) \rfloor$. Then

$$\begin{aligned} \frac{m_{i(n_l)}(F)}{i(n_l)!} &= \prod_{j=0}^{i(n_l)-1} m_1(F_j) = \prod_{j=0}^{i(n_l)-1} \frac{m_{j+1}(F)}{(j+1)m_j(F)} \\ &= \frac{m_{n_l}(F)}{n_l!} \prod_{j=n_l}^{i(n_l)-1} \frac{m_{j+1}(F)}{(j+1)m_j(F)} \geq \frac{m_{n_l}(F)}{n_l!} \left(\frac{1}{\lambda_0} + \frac{\eta}{2} \right)^{\lfloor n_l C \eta \rfloor} \end{aligned}$$

Now we need that

$$\liminf_{n \rightarrow \infty} \left(\frac{m_n(F)}{n!} \right)^{\frac{1}{n}} \geq \frac{1}{\lambda_0}. \quad (4.29)$$

Indeed, let $\epsilon > 0$. We know already that $\liminf_{n \rightarrow \infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} \geq \frac{1}{\lambda_0} > \frac{1}{\lambda_0 + \epsilon}$. So there exists $N \in \mathbb{N}$ such that for all $n > N$, $\frac{m_{n+1}(F)}{(n+1)m_n(F)} > \frac{1}{\lambda_0 + \epsilon}$ and thus $\frac{m_{n+1}(F)}{(n+1)} > \frac{m_n(F)}{\lambda_0 + \epsilon}$. Let Q be the constant, such that $\frac{m_{N+1}(F)}{(N+1)!} = \left(\frac{1}{\lambda_0 + \epsilon} \right)^{N+1} \cdot Q$. Then for all $n > N+1$ holds,

$$\frac{m_n(F)}{n!} \geq \frac{m_{n-1}(F)}{(n-1)!(\lambda_0 + \epsilon)} \geq \dots \geq \left(\frac{1}{\lambda_0 + \epsilon} \right)^{n-N-1} \frac{m_{N+1}(F)}{(N+1)!} = \left(\frac{1}{\lambda_0 + \epsilon} \right)^n \cdot Q$$

, so in total we got from this

$$\liminf_{n \rightarrow \infty} \left(\frac{m_n(F)}{n!} \right)^{\frac{1}{n}} \geq \frac{1}{\lambda_0 + \epsilon} \cdot \lim_{n \rightarrow \infty} Q^{\frac{1}{n}} = \frac{1}{\lambda_0 + \epsilon}.$$

Letting $\epsilon \rightarrow 0$ yields $\liminf_{n \rightarrow \infty} \left(\frac{m_n(F)}{n!} \right)^{\frac{1}{n}} \geq \frac{1}{\lambda_0}$.

4. The Main Theorem

So

$$\begin{aligned} \liminf_{l \rightarrow \infty} \left(\frac{m_{i(n_l)}(F)}{i(n_l)!} \right)^{\frac{1}{i(n_l)}} &\geq \liminf_{l \rightarrow \infty} \left[\left(\frac{m_{n_l}(F)}{n_l!} \right)^{\frac{1}{n_l}} \left(\frac{1}{\lambda_0} + \frac{\eta}{2} \right)^{C\eta} \right]^{\frac{n_l}{i(n_l)}} \\ &\geq \left[\frac{1}{\lambda_0} \left(\frac{1}{\lambda_0} + \frac{\eta}{2} \right)^{C\eta} \right]^{\frac{1}{1+C\eta}} > \frac{1}{\lambda_0}. \end{aligned}$$

But on the other hand, since $P_x(R > t) \leq Ce^{-\lambda_0 t}$ implies $m_k(F) \leq Ck! \left(\frac{1}{\lambda_0}\right)^k$ as in the proof of Theorem 3.7, we have

$$\liminf_{l \rightarrow \infty} \left(\frac{m_{i(n_l)}(F)}{i(n_l)!} \right)^{\frac{1}{i(n_l)}} \leq \lim_{l \rightarrow \infty} \left(\frac{Ci(n_l)! \left(\frac{1}{\lambda_0}\right)^{i(n_l)}}{i(n_l)!} \right)^{\frac{1}{i(n_l)}} = \frac{1}{\lambda_0}.$$

This is a contradiction. So we have shown that

$$\lim_{n \rightarrow \infty} \frac{m_{n+1}(F)}{(n+1)m_n(F)} = \frac{1}{\lambda_0}.$$

Lemma 3.5 implies now that F_n converges to an exponential distribution with mean $\frac{1}{\lambda_0}$. In addition,

$$E_{\Phi^n \delta_x}[R] = m_1(F^{\Phi^n \delta_x}) = m_1(F_n) = \frac{m_{n+1}(F)}{(n+1)m_n(F)} \rightarrow \frac{1}{\lambda_0}.$$

so we also proved (4.23). It only remains to show that

$$\lambda_0 = \left(\lim_{n \rightarrow \infty} E_{\Phi^n \delta_x} R \right)^{-1} < \infty.$$

We note that $\lim_{y \rightarrow \infty} E_y R = \infty$, because of our assumption $\lim_{y \rightarrow \infty} P_y(R \leq t) = 0$ for all $t \geq 0$ and also $E_y R > 0$ for all $y \in \mathbb{N}$. For shorter notation we set $\mu_n = \Phi^n \delta_x$. We assume for contradiction that $\lim_{n \rightarrow \infty} E_{\mu_n} R = 0$. For some fixed $S > 0$ there exists $Y \in \mathbb{N}$ such that for all $y > Y$ holds $E_y R \geq S$. Then $E_{\mu_n} R = \sum_{y \in \mathbb{N}} \mu_n(y) E_y R \geq \mu_n\{1, \dots, Y\} \min_{1 \leq y \leq Y} E_y R + S \cdot (1 - \mu_n\{1, \dots, Y\}) \rightarrow 0$ as $n \rightarrow \infty$. But this would imply $\mu_n\{1, \dots, Y\} \rightarrow 1$ and $\mu_n\{1, \dots, Y\} \rightarrow 0$, which is a contradiction. \square

4.6. Examples

Condition (4.3) is equivalent to the given condition in the following lemma, which is easier verified in examples.

4. The Main Theorem

Lemma 4.32. *Assume there exists a function $f : \mathbb{N}_0 \rightarrow \mathbb{R}_{\geq 0}$ and constants $D_1, D_4, D_5 > 0, D_2, D_3, D_6 < \infty, D_6 \in \mathbb{N}$ such that*

$$\lim_{x \rightarrow \infty} f(x) = \infty, \quad (4.30)$$

$$\sum_{y \neq x} \frac{q(x, y)}{q(x)} f(y) \leq f(x) - D_1 \text{ for } x \geq D_6, \quad (4.31)$$

$$|f(x) - f(y)| \leq D_2 \text{ for } x \geq D_6 \text{ and } q(x, y) > 0, \quad (4.32)$$

$$\sum_{y \neq x, f(y) \geq n} q(x, y) \leq D_3 e^{-D_4 n} \text{ for } 1 \leq x < D_6 \text{ and } n \geq 1, \quad (4.33)$$

and

$$-q(x, x) \geq D_5 \text{ for } x \geq D_6. \quad (4.34)$$

Then there exists $x, \lambda > 0$ such that $E_x e^{\lambda R} < \infty$.

Proof. Recall that J_1, J_2, \dots are the successive jump times of $(X(t))_{t \geq 0}$. For $\lambda \geq 0$ let

$$\Phi(\lambda) = [D_1 + D_2]^{-2} [e^{\lambda(D_1 + D_2)} - 1 - \lambda(D_1 + D_2)].$$

According to the results of Neveu [17] Lemma 7.2.8. on page 154, in a neighbourhood of 0 holds

$$\Phi(\lambda) = \frac{1}{2} \lambda^2 + o(\lambda^2) \quad (4.35)$$

and

$$e^{\lambda y} \leq 1 + \lambda y + \Phi(\lambda) y^2 \text{ for } y \leq (D_1 + D_2).$$

Now for $x \geq D_6$ holds almost surely on $\{X(0) = x\}$

$$\begin{aligned} & |f(X(J_1)) - f(X(0)) + D_1| \\ & \leq |f(X(J_1)) - f(X(0))| + D_1 \\ & \stackrel{(4.32)}{\leq} D_2 + D_1. \end{aligned}$$

So then

$$\begin{aligned} & E_x e^{\lambda[f(X(J_1)) - f(x) + D_1]} \\ & \leq E_x [1 + \lambda[f(X(J_1)) - f(x) + D_1] + \Phi(\lambda)[f(X(J_1)) - f(x) + D_1]^2] \\ & \stackrel{(4.32)}{\leq} 1 + E_x [\lambda[f(X(J_1)) - f(x) + D_1] + \Phi(\lambda)(D_1 + D_2)^2] \\ & \leq 1 + \Phi(\lambda)(D_1 + D_2)^2 \\ & \leq e^{\Phi(\lambda)(D_1 + D_2)^2}. \end{aligned}$$

4. The Main Theorem

where the third equality follows from

$$\begin{aligned} E_x f(X(J_1)) &= \sum_{y \neq x} \frac{q(x, y)}{q(x)} f(y) \\ &\stackrel{(4.31)}{\leq} (f(x) - D_1). \end{aligned}$$

For $0 < \epsilon < D_5$ follows then

$$\begin{aligned} &E_x e^{\lambda[f(X(J_1)) - f(x)] + \epsilon J_1} \\ &= E_x e^{\lambda[f(X(J_1)) - f(x)]} \cdot E_x e^{\epsilon J_1} \\ &= E_x e^{\lambda[f(X(J_1)) - f(x)]} \cdot \frac{q(x)}{q(x) - \epsilon} \\ &\leq e^{-\lambda D_1 + \Phi(\lambda)[D_1 + D_2]^2} \frac{D_5}{D_5 - \epsilon}. \end{aligned}$$

where $q(x) \geq D_5$ was used in the last inequality and where we used that the waiting time in a state and the next state are independent (see also Theorem 2.17). Because of (4.35), we can fix $\lambda > 0$ small enough such that

$$-\lambda D_1 + \Phi(\lambda)[D_1 + D_2]^2 \leq -\frac{\lambda D_1}{2},$$

and $\epsilon > 0$ such that

$$e^{-\lambda D_1 + \Phi(\lambda)[D_1 + D_2]^2} \frac{D_5}{D_5 - \epsilon} \leq 1.$$

With these choices, we define

$$g(t) = e^{\lambda[f(X(t)) - f(X(0))] + \epsilon t}$$

which satisfies

$$E_x g(J_1) \leq 1 = g(0). \quad (4.36)$$

Now we define $T := \inf\{t \geq J_1 : X(t) < D_6\} = J_{\inf\{k \geq 1 : X(J_k) < D_6\}}$ and $v := \inf\{k : J_k \geq T\} = \inf\{k \geq 1 : X(J_k) < D_6\}$, then $(Y_k)_{k \geq 0} = (g(J_{k \wedge v}))_{k \geq 0}$ is a positive supermartingale because of (4.36) under the measure \bar{P}_x for all $x \geq D_6$. It's positive, because g is positive. We have the filtration $(\mathcal{F}_k)_{k \in \mathbb{N}_0}$ given by $\mathcal{F}_k = \mathcal{F}_{J_k}$, i.e. the σ -algebra generated by the events up to the time of the k -th jump. We need to show for all $k \geq 0$ and all $x \geq D_6$ holds the inequality $E_x[Y_{k+1} | \mathcal{F}_k] \leq Y_k$. For $k = 0$, we have

$$E_x[Y_1 | \mathcal{F}_0] = E_x[Y_1 | \{\Omega, \emptyset\}] = E_x[Y_1] = E_x[g(J_{1 \wedge v})] = E_x[g(J_1)] \stackrel{(4.36)}{\leq} 1 = Y_0.$$

and for $k \geq 1$,

$$\begin{aligned} E_x[Y_{k+1} | \mathcal{F}_k] &= E_x[g(J_{k+1 \wedge v}) | \mathcal{F}_k] = g(J_{k \wedge v}) \mathbf{1}_{\{v \leq k\}} + E_x[g(J_{k+1}) | \mathcal{F}_k] \mathbf{1}_{\{v \geq k+1\}} \\ &= Y_k \mathbf{1}_{\{v \leq k\}} + E_x[e^{\lambda[f(X(J_k)) - f(X(0))] + \epsilon J_k} \cdot e^{\lambda[f(X(J_{k+1})) - f(X(J_k))] + \epsilon(J_{k+1} - J_k)} | \mathcal{F}_k] \\ &= Y_k \mathbf{1}_{\{v \leq k\}} + Y_k E_x[e^{\lambda[f(X(J_{k+1})) - f(X(J_k))] + \epsilon(J_{k+1} - J_k)} | \mathcal{F}_k] \mathbf{1}_{\{v \geq k+1\}} \\ &= Y_k \mathbf{1}_{\{v \leq k\}} + Y_k E_{X(J_k)}[e^{\lambda[f(X(J_1)) - f(X(0))] + \epsilon J_1}] \mathbf{1}_{\{v \geq k+1\}} \\ &\stackrel{(4.36)}{\leq} Y_k \mathbf{1}_{\{v \leq k\}} + Y_k \cdot 1 \cdot \mathbf{1}_{\{v \geq k+1\}} \leq Y_k, \end{aligned}$$

4. The Main Theorem

where we used the strong Markov property and also used in the next to last step that on $\{v \geq k+1\}$ we have $X(J_k) \geq D_6$.

Now since $Y_k \geq 0$ is a positive supermartingale, according to a Martingale convergence theorem (see for instance Durrett [9] Theorem 5.2.9. on page 236) there exists a random variable Y such that $Y_k \xrightarrow{k \rightarrow \infty} Y$ a.s. and $E[Y] \leq E[Y_0] = 1$. In our case clearly, the almost sure limit is $Y = g(J_v)$.

On the event $\{T \geq t\}$ we have $J_v \geq t$ and hence

$$g(J_v) = \exp(\lambda[f(X(J_v)) - f(X(0))] + \epsilon J_v) \stackrel{f \geq 0}{\geq} \exp(-\lambda f(X(0)) + \epsilon t).$$

Thus, for $x \geq D_6$,

$$\begin{aligned} P_x(T \geq t) &= E_x[\mathbf{1}_{T \geq t} \cdot \exp(-\lambda f(X(0)) + \epsilon t) \cdot \exp(\lambda f(x) - \epsilon t)] \\ &= E_x[\mathbf{1}_{T \geq t} \cdot \exp(-\lambda f(X(0)) + \epsilon t)] \exp(\lambda f(x) - \epsilon t) \\ &\leq E_x[\mathbf{1}_{T \geq t} g(J_v)] \exp(\lambda f(x) - \epsilon t) \\ &\leq E_x[g(J_v)] \exp(\lambda f(x) - \epsilon t) \\ &\leq \exp(\lambda f(x) - \epsilon t) \end{aligned}$$

where the last inequality follows from $(g(J_{k \wedge v}))_{k \geq 0}$ being a supermartingale and (4.36). This shows that T has an exponentially bounded tail, when $x \geq D_6$.

But then also for $x < D_6$, one gets with the strong Markov property

$$\begin{aligned} P_x(T \geq t) &= P_x(T \geq t, J_1 < t) + P_x(J_1 \geq t) \\ &= \sum_{y \geq D_6} P_y(T \geq t-s) P_x(X(J_1) = y, J_1 \in ds) + P_x(J_1 \geq t) \\ &\leq \sum_{y \geq D_6} \exp(\lambda f(y) - \epsilon(t-s)) P_x(X(J_1) = y, J_1 \in ds) + P_x(J_1 \geq t) \\ &\leq E_x(\exp(\lambda f(X(J_1)) + \epsilon J_1) \exp(-\epsilon t), X(J_1) \geq D_6) + P_x(J_1 \geq t) \\ &= \exp(-\epsilon t) E_x[e^{\epsilon J_1}] E_x[\exp(\lambda f(X(J_1))), X(J_1) \geq D_6] + P_x(J_1 \geq t) \\ &= \exp(-\epsilon t) \frac{q(x)}{q(x) - \epsilon} E_x[\exp(\lambda f(X(J_1))), X(J_1) \geq D_6] + e^{-q(x)t} \end{aligned}$$

Now if we choose $\lambda < D_4$, we have for all $n \in \mathbb{N}$ because of (4.33)

$$\sum_{y \geq D_6, f(y)=n} e^{\lambda f(y)} q(x, y) \leq e^{\lambda n} D_3 e^{-D_4 n} = D_3 e^{-(D_4 - \lambda)n}.$$

4. The Main Theorem

And hence we have

$$\begin{aligned}
E_x \left[e^{\lambda f(X(J_1))}, X(J_1) \geq D_6 \right] &= \sum_{y \geq D_6} e^{\lambda f(y)} \frac{q(x, y)}{q(x)} \\
&= \frac{1}{q(x)} \sum_{n \in \mathbb{N}} \sum_{y \geq D_6, f(y)=n} e^{\lambda f(y)} q(x, y) \\
&\leq \frac{D_3}{q(x)} \sum_{n \in \mathbb{N}} e^{-(D_4 - \lambda)n} < \infty.
\end{aligned}$$

So in total, for all $x \in \mathbb{N}$ there exists $\epsilon_x > 0$ small enough and a constant $K = K(x)$ such that

$$P_x(T \geq t) \leq K e^{-\epsilon t} \text{ for all } t \geq 0.$$

From this, it isn't difficult to follow that also R has exponentially bounded tails, and so because of Remark 4.9 there exists $\lambda > 0$ and $x \in \mathbb{N}$ such that $E_x e^{\lambda R} < \infty$. \square

This theorem shows, that our Main Theorem can be applied to any Markov process on the integers with bounded jumps and with a drift towards the origin.

4.6.1. The Random Walk

Here, we have a birth and death process with $\lambda_0 = \mu_0 = 0$ and $\lambda_i = \lambda$ and $\mu_i = \mu$ for all $i \in \mathbb{N}$ with $\mu > \lambda$. In Collet and Martinez [8] on page 105, it is demonstrated that there are infinitely many quasi-stationary distributions.

4.6.2. Linear birth and death process

We have a birth and death process with $\mu_0 = \lambda_0 = 0$ and we are given $0 < \lambda < \mu$ and $\lambda_n = n\lambda, \mu_n = n\mu$ to be the birth and death coefficients. Before we check that there exists an exponential moment of R , we see that this chain is not absorbed in finite time when coming from infinity, i.e. we check that for all $t > 0$ we have

$$\lim_{x \rightarrow \infty} P_x(R \leq t) = 0.$$

Indeed let $t > 0$ be fixed. Then

$$\begin{aligned}
P_x(R \leq t) &\leq P_x(X(t) = 0 | X(J_1) = x - 1, \dots, X(J_x) = 0) \\
&= P_x(T_0 + \dots + T_{x-1} \leq t | X_1 = x - 1, \dots, X_x = 0) \\
&\leq P_x(G \leq t)
\end{aligned}$$

where G is the convolution of x independent exponentially distributed random variables with parameters $x(\lambda + \mu)$, i.e. G is gamma-distributed with parameters $x(\lambda + \mu)$ and x and has the density

$$f_G(k) = \frac{(x(\lambda + \mu))^x}{(x-1)!} e^{-x(\lambda + \mu)k} k^{x-1} \mathbf{1}_{\mathbb{R}_{\geq 0}}(k).$$

4. The Main Theorem

So continuing the calculation, we have

$$\begin{aligned} P_x(R \leq t) &\leq \int_0^t \frac{(x(\lambda + \mu))^x}{(x-1)!} e^{-x(\lambda + \mu)k} k^{x-1} dk \\ &= \int_{x(\lambda + \mu)t}^{\infty} k^{x-1} e^{-k} dk / (x-1)! \xrightarrow{x \rightarrow \infty} 0. \end{aligned}$$

Now we show the existence of a quasi-stationary distribution by verifying the conditions of Lemma 4.32. If we choose $f(x) = x$, $D_6 = 1$, $D_1 = \frac{\mu - \lambda}{\lambda + \mu} > 0$, $D_2 = 1$, $D_5 = \lambda + \mu$, then obviously all conditions except (4.31) hold. But also (4.31) holds, because for all $x \geq D_6 = 1$ we have

$$\begin{aligned} E_x[f(X(J_1))] &= \sum_{y \neq x} \frac{q(x, y)}{q(x)} f(y) \\ &= \frac{q(x, x-1)}{q(x)} f(x-1) + \frac{q(x, x+1)}{q(x)} f(x+1) \\ &= \frac{x\mu}{x(\lambda + \mu)} (x-1) + \frac{x\lambda}{x(\lambda + \mu)} (x+1) \\ &= x + \frac{\lambda - \mu}{\lambda + \mu} = x - D_1 = f(x) - D_1. \end{aligned}$$

4.6.3. M/M/ ∞ Queue

In this case, we have a birth and death process with $\lambda_0 = \mu_0 = 0$ and $\lambda_i = \lambda$ and $\mu_i = i\mu$ for $i \in \mathbb{N}$. Again, we verify the conditions of Lemma 4.32 in order to show that a quasi-stationary distribution exists.

We choose $f(x) = x$, $D_1 = \frac{1}{2} > 0$, $D_2 = 1$, $D_4 = 1$, $D_5 = \mu + \lambda$ and D_6 such that for all $x \geq D_6$ we have $\frac{\lambda - x\mu}{x\mu + \lambda} < -\frac{1}{2}$ and D_3 such that $(\lceil 3\frac{\lambda}{\mu} \rceil - 1)\mu + \lambda \leq D_3 e^{-\lceil 3\frac{\lambda}{\mu} \rceil}$, then obviously all conditions from Lemma 4.32 except (4.31) and (4.33) hold. But also (4.31) holds because for all $x \geq D_6$ holds

4. The Main Theorem

$$\begin{aligned}
E_x[f(X(J_1))] &= \sum_{y \neq x} \frac{q(x, y)}{q(x)} \\
&= \frac{q(x, x-1)}{q(x)} f(x-1) + \frac{q(x, x+1)}{q(x)} f(x+1) \\
&= \frac{x\mu}{x\mu + \lambda} (x-1) + \frac{\lambda}{x\mu + \lambda} (x+1) \\
&= \frac{x(x\mu + \lambda)}{x\mu + \lambda} + \frac{\lambda - x\mu}{x\mu + \lambda} \\
&= x + \frac{\lambda - x\mu}{x\mu + \lambda} \leq x - \frac{1}{2} = f(x) - D_1.
\end{aligned}$$

Now we need to check (4.33). By our choice of D_6 for all $x \geq D_6$ holds

$$\frac{\lambda - x\mu}{x\mu + \lambda} < -\frac{1}{2} \iff \lambda - x\mu < \left(-\frac{1}{2}\right)(x\mu + \lambda) \iff \frac{3}{2}\lambda < \frac{1}{2}x\mu \iff 3\frac{\lambda}{\mu} < x.$$

So we can assume that $D_6 = \lceil 3\frac{\lambda}{\mu} \rceil$. Thus for all $1 \leq x < D_6$ holds:

If $n > D_6$ then trivially

$$\sum_{y \neq x, f(y) \geq n} q(x, y) = 0 \leq D_3 e^{-D_4 n}.$$

For the interesting case $n \leq D_6$ we have

$$\begin{aligned}
\sum_{y \neq x, f(y) \geq n} q(x, y) &\leq q(x, x-1) + q(x, x+1) \\
&= x\mu + \lambda \\
&\leq \left(\lceil 3\frac{\lambda}{\mu} \rceil - 1\right)\mu + \lambda \\
&\leq D_3 e^{-\lceil 3\frac{\lambda}{\mu} \rceil} \\
&= D_3 e^{-D_4 \lceil 3\frac{\lambda}{\mu} \rceil} \leq D_3 e^{-D_4 n}.
\end{aligned}$$

4.6.4. Markov branching process

We remind here Definition 2.28 of a continuous-time Markov branching process. We now show the existence of a quasi-stationary distribution for a Markov branching process, which has basic generating function given by the parameters

$$p_0 = \frac{7}{10}, p_1 = p_2 = p_3 = \frac{1}{10}.$$

5. Minimal quasi-stationary distributions

We show the existence of a quasi-stationary distribution by verifying the conditions of Lemma 4.32. If we choose $f(x) = x$, $D_6 = 1$, $D_1 = -\frac{p_2+2p_3-p_0}{1-p_1} > 0$, $D_2 = 2$, $D_5 = a(1-p_1)$, then obviously all conditions except (4.31) hold. But also (4.31) holds, because for all $x \geq D_6 = 1$ we have

$$\begin{aligned} E_x[f(X(J_1))] &= \sum_{y \neq x} \frac{q(x, y)}{q(x)} f(y) \\ &= \frac{q(x, x-1)}{q(x)} f(x-1) + \frac{q(x, x+1)}{q(x)} f(x+1) + \frac{q(x, x+2)}{q(x)} f(x+2) \\ &= \frac{xa p_0}{xa(1-p_1)} (x-1) + \frac{xa p_2}{xa(1-p_1)} (x+1) + \frac{xa p_3}{xa(1-p_1)} (x+2) \\ &= \frac{1}{1-p_1} (x(p_0+p_2+p_3) + p_2 + 2p_3 - p_0) = x + \frac{p_2+2p_3-p_0}{1-p_1} = f(x) - D_1. \end{aligned}$$

5. Minimal quasi-stationary distributions

5.1. Setting and Main Theorem

Define for a probability distribution μ on \mathbb{N} ,

$$\lambda(\mu) = \sup\{\lambda : E_\mu e^{\lambda R} < \infty\}. \quad (5.1)$$

Let $x \in \mathbb{N}$ in the support of μ , then for all λ with $E_\mu e^{\lambda R} < \infty$ holds $\infty > \frac{E_\mu e^{\lambda R}}{\mu(x)} \geq E_x e^{\lambda R}$ and thus, we have $\lambda(\mu) \leq \lambda_0$.

If μ is a quasi-stationary distribution, then we know from the proof of the easy direction of Theorem 4.6, that F^μ is exponential with mean $E_\mu R < \infty$ and that $\lambda(\mu) = \frac{1}{E_\mu R} \leq \lambda_0$. If there is equality, we call μ a **minimal quasi-stationary distribution**, because the mean time to absorption is minimal.

In subsection 4.4, we showed under the hypothesis $\lim_{x \rightarrow \infty} P_x(R < t) = 0$ that for any $\theta > 0$ for which \mathcal{M}_θ is non-empty, \mathcal{M}_θ contains a quasi-stationary distribution. In this section, we show that we can get a minimal quasi-stationary distribution corresponding to $\theta = \frac{1}{\lambda_0}$ as sub-sequential limit of $\Phi^n \delta_x$, when we impose on Q the additional assumption

$$q(x) = -q(x, x) = \sum_{y \neq x} q(x, y) \leq C_1 < \infty, \text{ for some constant } C_1 \text{ and all } x \in \mathbb{N}. \quad (5.2)$$

The big result of this section is:

Theorem 5.1. *Assume (5.2) and (4.3) hold.*

5. Minimal quasi-stationary distributions

- (a) If $(n_k)_{k \in \mathbb{N}}$ is a subsequence such that for some $x \in \mathbb{N}$ we have that $\Phi^{n_k} \delta_x$ converges weakly to a probability distribution μ_∞ on \mathbb{N} , then μ_∞ is a minimal quasi-stationary distribution.
- (b) Similarly, if $(t_k)_{k \in \mathbb{N}}$ is a sequence of positive real numbers with $t_k \rightarrow \infty$ and μ_∞ is a probability distribution on \mathbb{N} such that for some $x \in \mathbb{N}$,

$$\lim_{k \rightarrow \infty} \frac{P_x(X(t_k) = y)}{P_x(R > t_k)} = \mu_\infty(y), y \in \mathbb{N}, \quad (5.3)$$

then μ_∞ is a minimal quasi-stationary distribution.

For the proof, we need some results about discrete Markov chains that are taken from Kesten [13].

Definition 5.2. Let $P = P(i, j)_{i, j \in \mathbb{N}}$ be a matrix. It is called **positive** if $P(i, j) \geq 0$ for all $i, j \in \mathbb{N}$. It is called **irreducible**, if for all $i, j \in \mathbb{N}$ there exists $n = n(i, j) > 0$ such that $P^{(n)}(i, j) > 0$. It is called **aperiodic** if for all $i \in \mathbb{N}$, $\gcd\{n : P^{(n)}(i, i) > 0\} = 1$. And finally it is called **sub-stochastic** if it is positive and for all $i \in \mathbb{N}$, $\sum_{j \in \mathbb{N}} P(i, j) \leq 1$.

Lemma 5.3. Let P be a positive, irreducible and aperiodic matrix. Then the limit

$$\lim_{n \rightarrow \infty} [P^{(n)}(i, j)]^{\frac{1}{n}} = R^{-1} \in [0, \infty] \quad (5.4)$$

exists for all i, j and is independent of i, j .

Proof. See Kesten [13] Lemma 3 on page 661. □

Lemma 5.4. Let P be a sub-stochastic, irreducible matrix. Further assume that P is **uniformly aperiodic**, i.e. assume there exists constants $\delta_1 > 0$ and $N < \infty$ such that for each $i \in \mathbb{N}$ there are integers $1 \leq k_1, \dots, k_r \leq N$ (with $k_j = k_j(i)$ and $r = r(i)$) such that $P^{(k_s)}(i, i) \geq \delta_1$ for $1 \leq s \leq r$ and $\gcd(k_1, \dots, k_r) = 1$. Then with R as in (5.4), it is $1 \leq R < \infty$ (so $1 \geq \frac{1}{R} > 0$) and for all $i, j, l \in \mathbb{N}$ holds

$$\lim_{n \rightarrow \infty} \frac{P^{(n+l)}(i, j)}{P^{(n)}(i, j)} = R^{-l}. \quad (5.5)$$

Proof. See Kesten [13] Lemma 4 on page 661. □

Remark 5.5. For every $h > 0$, the matrix P_h on \mathbb{N} given by $P_h(x, y) = P_x(X(h) = y)$ is a substochastic matrix, that is irreducible and aperiodic because of Proposition 2.6. Moreover with $\delta_1 = e^{-C_1 h}$, we have for all $x \in \mathbb{N}$,

$$\begin{aligned} P_h(x, x) &= P_x(X(h) = x) \geq P_x(T_0 > h) \\ &= e^{-q(x)h} \stackrel{(5.2)}{\geq} e^{-C_1 h} = \delta_1. \end{aligned}$$

So P_h satisfies the assumptions from Lemmas 5.3 and 5.4.

5. Minimal quasi-stationary distributions

Lemma 5.6. *If (5.2) holds, then there exists $\lambda_1 \geq 0$ such that for all $h > 0$*

$$\lim_{k \rightarrow \infty} (P_x(X(hk) = y))^{\frac{1}{k}} = e^{-\lambda_1 h}, \quad (5.6)$$

and for all $l \geq 1$,

$$\lim_{k \rightarrow \infty} \frac{P_x(X(h(k+l)) = y)}{P_x(X(hk) = y)} = e^{-\lambda_1 hl}. \quad (5.7)$$

Proof. For each $h > 0$ the matrix P_h given by $P_h(x, y) = P_x(X(h) = y)$ satisfies the conditions from Lemma 5.4. Thus there is $1 \leq R_h < \infty$ such that $\lim_{k \rightarrow \infty} \left(P_h^{(k)}(x, y)\right)^{\frac{1}{k}} = R_h^{-1}$ for all $x, y \in \mathbb{N}$. Let $\lambda_h \geq 0$ be such that $e^{-\lambda_h h} = R_h^{-1}$. Then we have for all $h > 0$

$$\lim_{k \rightarrow \infty} (P_x(X(kh) = y))^{\frac{1}{kh}} = e^{-\lambda_h}.$$

Because $(P_{xy}(t))^{\frac{1}{t}}$ is continuous and Proposition A.4, λ_h is independent of h and with λ_1 the common value of the λ_h , we got (5.6). With (5.6) and Lemma 5.4, we get (5.7). \square

Lemma 5.7. *If (5.2) holds, then for some $\lambda_1 \in [\lambda_0, \infty)$ we have for all $x, y \in \mathbb{N}$,*

$$\lim_{t \rightarrow \infty} (P_x(X(t) = y))^{\frac{1}{t}} = e^{-\lambda_1}. \quad (5.8)$$

Moreover, for all $x, y \in \mathbb{N}$ and all $s \geq 0$,

$$\lim_{t \rightarrow \infty} \frac{P_x(X(t+s) = y)}{P_x(X(t) = y)} = e^{-\lambda_1 s}. \quad (5.9)$$

(λ_1 is independent of x and y .)

Proof. We know from Lemma 5.6 that

$$\lim_{k \rightarrow \infty} (P_x(X(hk) = y))^{\frac{1}{kh}} = e^{-\lambda_1} \quad (5.10)$$

for all $x, y \in \mathbb{N}, h > 0$.

Now with $k \in \mathbb{N}$ such that $k \leq t \leq k+1$, we have on the one hand

$$\begin{aligned} P_x(X(t) = y)^{\frac{1}{t}} &\geq (P_x(X(t-k) = x) \cdot P_x(X(k) = y))^{\frac{1}{k}} \\ &\geq \underbrace{\left(\min_{u \in [0,1]} P_x(X(u) = x)\right)^{\frac{1}{k}}}_{\xrightarrow[t \rightarrow \infty]{} 1} P_x(X(k) = y)^{\frac{1}{k}} \xrightarrow[t \rightarrow \infty]{} e^{-\lambda_1}. \end{aligned}$$

5. Minimal quasi-stationary distributions

and on the other hand

$$\begin{aligned} P_x(X(t) = y)^{\frac{1}{t}} &\leq (P_x(X(k+1) = y) \cdot P_y(X(k+1-t) = y)^{-1})^{\frac{1}{k+1}} \\ &\leq P_x(X(k+1) = y)^{\frac{1}{k+1}} \left(\min_{u \in [0,1]} P_y(X(u) = y) \right)^{-\frac{1}{k+1}} \xrightarrow{t \rightarrow \infty} e^{-\lambda_1}. \end{aligned}$$

and this shows (5.8).

For (5.9), let $s \geq 0$. Let k be such that $kh \leq t \leq (k+1)h$ and l be such that $(l-1)h \leq s \leq lh$. Then holds

$$\begin{aligned} \frac{P_x(X(t+s) = y)}{P_x(X(t) = y)} &\leq \frac{P_x(X(h(k+1+l)) = y) P_y(X(h(k+1+l) - (t+s)) = y)^{-1}}{P_x(X(hk) = y) \cdot P_y(X(t-hk) = y)} \\ &\leq \frac{P_x(X(h(k+1+l)) = y)}{P_x(X(hk) = y)} \cdot \min_{u \in [0,2h]} P_y(X(u) = y)^{-2} \\ &\xrightarrow{t \rightarrow \infty} e^{-\lambda_1(l+1)h} \min_{u \in [0,2h]} P_y(X(u) = y)^{-2} \xrightarrow{h \rightarrow 0} e^{-\lambda_1 s}. \end{aligned}$$

and on the other hand

$$\begin{aligned} \frac{P_x(X(t+s) = y)}{P_x(X(t) = y)} &\geq \frac{P_x(X(h(k+l-1)) = y) P_y(X(t+s-h(k+l-1)) = y)}{P_x(X(h(k+1)) = y) \cdot P_y(X(h(k+1)-t) = y)^{-1}} \\ &\geq \frac{P_x(X(h(k+l-1)) = y)}{P_x(X(h(k+1)) = y)} \cdot \min_{u \in [0,2h]} P_y(X(u) = y)^2 \\ &\xrightarrow{t \rightarrow \infty} e^{-\lambda_1 h(l-2)} \min_{u \in [0,2h]} P_y(X(u) = y)^2 \xrightarrow{h \rightarrow 0} e^{-\lambda_1 s}. \end{aligned}$$

and this shows (5.9).

It remains to show that $\lambda_1 \geq \lambda_0$. By the definition of λ_0 , for all $\epsilon > 0$ and t sufficiently large, we have

$$\begin{aligned} P_x(R > t) &= P_x\left(e^{(\lambda_0 - \frac{\epsilon}{2})R} > e^{(\lambda_0 - \frac{\epsilon}{2})t}\right) \\ &\leq E_x\left[e^{(\lambda_0 - \frac{\epsilon}{2})R}\right] e^{-(\lambda_0 - \frac{\epsilon}{2})t} \leq e^{\frac{\epsilon}{2}t} e^{-(\lambda_0 - \epsilon/2)t} \\ &\leq e^{-(\lambda_0 - \epsilon)t}. \end{aligned}$$

But on the other hand, because of (5.8) for all $\epsilon > 0$ and t large enough, we have

$$e^{-(\lambda_1 + \epsilon)} \leq P_x(X(t) = x)^{1/t}.$$

Thus, we have for all $\epsilon > 0$ and t sufficiently large

$$e^{-(\lambda_1 + \epsilon)} \leq P_x(X(t) = x)^{1/t} \leq \left(\sum_{y \in \mathbb{N}} P_x(X(t) = y) \right)^{1/t} = P_x(R > t)^{1/t} \leq e^{-(\lambda_0 - \epsilon)}.$$

5. Minimal quasi-stationary distributions

And this shows that it must be $\lambda_1 \geq \lambda_0$. □

Proof of Theorem 5.1. By Theorem 4.13, for any probability measure ν on \mathbb{N} with $E_\nu R < \infty$ and $y \in \mathbb{N}$ holds

$$\Phi \nu(y) = \frac{1}{E_\nu R} \int_0^\infty \sum_{x \in \mathbb{N}} \nu(x) P_x(X(t) = y) dt.$$

Now, we show by induction over n , that for all $n \in \mathbb{N}$ holds

$$\begin{aligned} \Phi^n \nu(y) &= \prod_{k=0}^{n-1} (E_{\Phi^k \nu} R)^{-1} \int_0^\infty \dots \int_0^\infty P_\nu(X(t_1 + \dots + t_n) = y) dt_n \dots dt_1 \\ &= \prod_{k=0}^{n-1} (E_{\Phi^k \nu} R)^{-1} \frac{1}{(n-1)!} \int_0^\infty t^{n-1} P_\nu(X(t) = y) dt. \end{aligned} \quad (5.11)$$

This is clear for $n = 1$. Assume now for $n \in \mathbb{N}$ (5.11) holds. Then

$$\begin{aligned} \Phi^{n+1} \nu(y) &= \frac{1}{E_{\Phi^n \nu} R} \int_0^\infty \sum_{x \in \mathbb{N}} \Phi^n \nu(x) P_x(X(t) = y) dt \\ &= \frac{1}{E_{\Phi^n \nu} R} \int_0^\infty \sum_{x \in \mathbb{N}} \prod_{k=0}^{n-1} (E_{\Phi^k \nu} R)^{-1} \int_0^\infty \dots \int_0^\infty P_\nu(X(t_1 + \dots + t_n) = x) dt_n \dots dt_1 \\ &\quad \cdot P_x(X(t) = y) dt \\ &= \prod_{k=0}^n (E_{\Phi^k \nu} R)^{-1} \int_0^\infty \dots \int_0^\infty \sum_{x \in \mathbb{N}} P_\nu(X(t_1 + \dots + t_n) = x) \cdot \\ &\quad \cdot P(X(t_1 + \dots + t_{n+1}) = y | X(t_1 + \dots + t_n) = x) dt_{n+1} \dots dt_1 \\ &= \prod_{k=0}^n (E_{\Phi^k \nu} R)^{-1} \int_0^\infty \dots \int_0^\infty P_\nu(X(t_1 + \dots + t_{n+1}) = y) dt_{n+1} \dots dt_1. \end{aligned}$$

If we fix $x \in \mathbb{N}$ and write

$$c_n = \frac{1}{(n-1)!} \prod_{k=0}^{n-1} (E_{\Phi^k \delta_x} R)^{-1},$$

then

$$(\Phi^n \delta_x)(y) = c_n \int_0^\infty t^{n-1} P_x(X(t) = y) dt.$$

Assume now there is a subsequence $(n_k)_{k \in \mathbb{N}}$ such that $\Phi^{n_k} \delta_x \Rightarrow \mu_\infty$ for some probability

5. Minimal quasi-stationary distributions

measure μ_∞ on \mathbb{N} . Then for $s \geq 0, y \in \mathbb{N}$,

$$\begin{aligned}
P_{\mu_\infty}(X(s) = y) &= \sum_{z \in \mathbb{N}} \mu_\infty(z) P_z(X(s) = y) \\
&= \lim_{Z \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{z=1}^Z \Phi^{n_k} \delta_x(z) P_z(X(s) = y) \\
&= \lim_{k \rightarrow \infty} \sum_{z \in \mathbb{N}} \Phi^{n_k} \delta_x(z) P_z(X(s) = y) \\
&= \lim_{k \rightarrow \infty} \sum_{z \in \mathbb{N}} c_{n_k} \int_0^\infty t^{n_k-1} P_x(X(t) = z) P(X(t+s) = y | X(t) = z) dt \\
&= \lim_{k \rightarrow \infty} c_{n_k} \int_0^\infty t^{n_k-1} P_x(X(t+s) = y) dt. \tag{5.12}
\end{aligned}$$

We note here that after the Theorem of Prohorov, because the family $(\Phi^{n_k} \delta_x)_{k \in \mathbb{N}}$ is weakly relative sequentially compact, it is tight. And it is therefore that we can exchange the limits in the second step above as we did in (4.19).

If we can show, that for each fixed T the contribution of the integral over $[0, T]$ in (5.12) is negligible, then it follows from (5.9) that

$$P_{\mu_\infty}(X(s) = y) = \lim_{k \rightarrow \infty} e^{-\lambda_1 s} c_{n_k} \int_0^\infty t^{n_k-1} P_x(X(t) = y) dt = e^{-\lambda_1 s} \mu_\infty(y),$$

and by summing over $y \in \mathbb{N}$,

$$P_{\mu_\infty}(R > s) = \sum_{y \in \mathbb{N}} P_{\mu_\infty}(X(s) = y) = e^{-\lambda_1 s}. \tag{5.13}$$

and hence

$$P_{\mu_\infty}(X(s) = y | R > s) = \mu_\infty(y).$$

and μ_∞ is a quasi-stationary distribution.

μ_∞ is also minimal, because on the one hand $\lambda_1 \geq \lambda_0$ and on the other hand we have for $\epsilon > 0$,

$$\begin{aligned}
E_{\mu_\infty}[e^{(\lambda_1 - \epsilon)R}] &= 1 + \int_0^\infty P_{\mu_\infty}(e^{(\lambda_1 - \epsilon)R} - 1 > t) dt \\
&= 1 + \int_0^\infty P_{\mu_\infty}\left(R > \frac{\ln(t+1)}{\lambda_1 - \epsilon}\right) dt \\
&\stackrel{(5.13)}{=} 1 + \int_1^\infty e^{-\frac{\lambda_1}{\lambda_1 - \epsilon} \ln(t)} dt \\
&= 1 + \int_1^\infty t^{-\frac{\lambda_1}{\lambda_1 - \epsilon}} dt < \infty.
\end{aligned}$$

5. Minimal quasi-stationary distributions

Hence, we have for every x in the support of μ_∞ and every $\epsilon > 0$

$$E_x [e^{(\lambda_1 - \epsilon)R}] \leq \frac{E_{\mu_\infty} [e^{(\lambda_1 - \epsilon)R}]}{\mu_\infty(x)} < \infty.$$

which shows that $\lambda_1 = \lambda_0$ and μ_∞ a minimal quasi-stationary distribution.

It remains to show that for all T the integral of (5.12) over $[0, T]$ can be neglected. On the one hand,

$$\int_0^T t^{n-1} P_x(X(t) = y) dt \leq \int_0^T t^{n-1} dt = \frac{T^n}{n}. \quad (5.14)$$

while on the other hand, for $0 < \epsilon < \frac{\lambda_1}{2}$ and n large, we have

$$\begin{aligned} \int_0^\infty t^{n-1} P_x(X(t) = y) dt &\geq \int_{\frac{n}{2e\lambda_1}}^\infty t^{n-1} e^{-(\lambda_1 + \epsilon)t} dt \quad [\text{by (5.8)}] \\ &\geq \int_0^\infty t^{n-1} e^{-(\lambda_1 + \epsilon)t} dt - \frac{n^{n-1}}{(2e\lambda_1)^n} \\ &= \frac{(n-1)!}{(\lambda_1 + \epsilon)^n} - \frac{n^{n-1}}{(2e\lambda_1)^n} \\ &= \frac{(n-1)!}{2(\lambda_1 + \epsilon)^n} + \frac{(n-1)!}{2(\lambda_1 + \epsilon)^n} - \frac{n^{n-1}}{(2e\lambda_1)^n} \\ &\geq \frac{(n-1)!}{2(\lambda_1 + \epsilon)^n}. \end{aligned}$$

The last inequality follows from the underneath calculation:

$$\begin{aligned} \frac{(n-1)!}{2(\lambda_1 + \epsilon)^n} - \frac{n^{n-1}}{(2e\lambda_1)^n} &\geq \frac{n!}{2n(2\lambda_1)^n} - \frac{2n^n}{2n(2\lambda_1)^n e^n} \quad [\text{because } \epsilon < \lambda_1/2] \\ &= \frac{1}{2n(2\lambda_1)^n} \left(n! - 2 \left(\frac{n}{e} \right)^n \right) \geq 0, \end{aligned}$$

where the last inequality, we prove by induction over n . For $n = 1$ this is clear. And then we note that $(1 + \frac{1}{n})^n$ converges monotone increasingly to e , then we get

$$\begin{aligned} (n+1)! &= n! \cdot (n+1) > 2 \left(\frac{n}{e} \right)^n (n+1) \\ &= 2 \left(\frac{n+1}{e} \right)^{n+1} \cdot \left(\frac{n}{n+1} \right)^n \cdot e = 2 \left(\frac{n+1}{e} \right)^{n+1} \frac{1}{(1 + \frac{1}{n})^n} \cdot e > 2 \left(\frac{n+1}{e} \right)^{n+1}. \end{aligned}$$

Therefore, we have

$$\int_0^\infty t^{n-1} P_x(X(t) = y) dt \geq \frac{(n-1)!}{2(\lambda_1 + \epsilon)^n}. \quad (5.15)$$

5. Minimal quasi-stationary distributions

Because the contribution of the integral over $[0, T]$ does increase at most as fast as $\frac{T^n}{n}$, this shows that for large n this contribution is negligible and this concludes part one of the Theorem.

Now if μ_∞ is such that (5.3) holds. Then

$$\begin{aligned} P_{\mu_\infty}(X(s) = y) &= \sum_{z \in \mathbb{N}} \mu_\infty(z) P_z(X(s) = y) \\ &= \lim_{Z \rightarrow \infty} \lim_{k \rightarrow \infty} \sum_{z=1}^Z \frac{P_x(X(t_k) = z)}{P_x(R > t_k)} \cdot P_z(X(s) = y) \\ &= \lim_{k \rightarrow \infty} \sum_{z \in \mathbb{N}} c_k P_x(X(t_k) = z) P_z(X(s) = y), \end{aligned}$$

where

$$c_k := (P_x(R > t_k))^{-1}.$$

We note again, that the family $(P_k)_{k \in \mathbb{N}} = (P_x(X(t_k) = \bullet | R > t_k))_{k \in \mathbb{N}}$ is weakly relative sequentially compact, and thus after Prohorov's theorem, it is tight, so we can exchange limits above with the same argument as in (4.19). Continuing, we have

$$\begin{aligned} P_{\mu_\infty}(X(s) = y) &= \lim_{k \rightarrow \infty} c_k \sum_{z \in \mathbb{N}} P_x(X(t_k) = z) P(X(t_k + s) = y | X(t_k) = z) \\ &= \lim_{k \rightarrow \infty} c_k P_x(X(t_k + s) = y) \\ &= e^{-\lambda_1 s} \lim_{k \rightarrow \infty} c_k P_x(X(t_k) = y) \text{ [by (5.9)]} \\ &= e^{-\lambda_1 s} \mu_\infty(y). \end{aligned}$$

This is (5.13) again, and we prove μ_∞ is a minimal quasi-stationary distribution as before. □

5.2. Birth and Death Process

Here, we treat birth and death processes with 0 as absorbing state, i.e. $q(0, y) = 0$ for all $y \in \mathbb{N}_0$. Further we assume, that the holding parameters $q(x)$ are uniformly bounded, i.e. there is $C < \infty$ such that

$$q(x) = -q(x, x) < C, \quad \text{for all } x \in \mathbb{N}_0. \quad (5.16)$$

In this case, the minimal Q -function $P_{xy}(t)$ is the only Q -function and is honest (see [2] Proposition 2.10 on page 84) and the minimal process $X = (X(t))_{t \geq 0}$ can be constructed as a Markov jump process.

5. Minimal quasi-stationary distributions

In this setting, also (4.2) is satisfied. Indeed, we have

$$\begin{aligned} P_x(R > t) &= P_x(X(t) \neq 0) \\ &= E_x[P_x(X(t) \neq 0 | X_0, X_1, \dots, X_{x-1})] \\ &\geq E_x[P_x(T_0 + T_1 + \dots + T_{x-1} > t | X_0, \dots, X_{x-1})] \end{aligned}$$

because we need at least x jumps to reach 0 starting at $x \in \mathbb{N}$. Given X_0, \dots, X_{x-1} Theorem 2.17 tells that T_0, \dots, T_{x-1} are independent with T_i being exponential with parameter $q(X_i) \leq C$. Thus if W_0, W_1, \dots, W_{x-1} are independent identically distributed with W_0 being exponentially distributed with parameter C , then

$$P_x(T_0 + T_1 + \dots + T_{x-1} > t | X_1, \dots, X_x) \geq P(W_0 + \dots + W_{x-1} > t) \xrightarrow{x \rightarrow \infty} 1.$$

because of the strong law of large numbers. So $\lim_{x \rightarrow \infty} P_x(R > t) = 1$.

We further assume, that all neighbouring, non-zero states connect, i.e. for all $x \in \mathbb{N}$

$$q(x, x+1) = \lambda_x > 0 \text{ and } q(x, x-1) = \mu_x > 0. \quad (5.17)$$

and that there is $\lambda > 0$ such that $E_x e^{\lambda R} < \infty$. Then by Theorem 4.6, there exists a quasi-stationary distribution.

Theorem 5.8. $\lim_{n \rightarrow \infty} \Phi^n \delta_x$ exists and is the unique minimal quasi-stationary distribution.

First, we prove that the distribution of the absorption time determines the initial measure.

Lemma 5.9. *The function $v \mapsto F^v$ is injective.*

Proof. Before we can prove this lemma, we need the following generalization of the Kolmogorov backward equations.

Lemma 5.10. *For all $n \in \mathbb{N}, x, y \in \mathbb{N}_0, t \geq 0$ we have*

$$\frac{d^n}{dt^n} P_{xy}(t) = \sum_{z=0}^{\infty} q^{(n)}(x, z) P_{zy}(t) \quad (5.18)$$

where P is the transition function of X and $q^{(n)}(x, z)$ denotes the (x, z) entry of the matrix Q^n .

Proof. We prove this by induction over n .

5. Minimal quasi-stationary distributions

$\mathbf{n} = \mathbf{1}$:

Because Q is conservative, the transition function $P_{ij}(t)$ of X satisfies the Kolmogorov backward equations

$$P'_{xy}(t) = \sum_{z=0}^{\infty} q(x, z) P_{zy}(t) \text{ for all } t \geq 0 \text{ and } j \in E.$$

So let 5.18 hold for some $n \in \mathbb{N}$.

$\mathbf{n} \rightarrow \mathbf{n} + \mathbf{1}$:

$$\begin{aligned} \frac{d^{n+1}}{dt^{n+1}} P_{xy}(t) &= \frac{d}{dt} \left(\frac{d^n}{dt^n} P_{xy}(t) \right) \\ &= \frac{d}{dt} \left(\sum_{z=0}^{\infty} q^{(n)}(x, z) P_{zy}(t) \right). \end{aligned}$$

Now in order to exchange the sum and the differentiation, we need to apply theorem A.6. So we need to show that

$$\sum_{z=0}^{\infty} \left\| \frac{d}{dt} (q^{(n)}(x, z) P_{zy}(t)) \right\| < \infty. \quad (5.19)$$

For this we have at our disposal, that the holding time parameters $q(x) \leq C, x \in \mathbb{N}_0$ are uniformly bounded, that $q^{(n)}(x, z) = 0$ if $|x - z| > n$ and that $\sum_{y \in \mathbb{N}_0} \left| \frac{d}{dt} P_{zy}(t) \right| \leq 2q(z) \leq 2C$ for all $z \in \mathbb{N}_0, t \geq 0$ (see Anderson [2] Proposition 2.6. on page 12). Thus

$$\sum_{z=0}^{\infty} \left\| \frac{d}{dt} q^{(n)}(x, z) P_{zy}(t) \right\| = \sum_{z=\max\{0, x-n\}}^{x+n} q^{(n)}(x, z) \underbrace{\left\| \frac{d}{dt} P_{zy}(t) \right\|}_{\leq 2C} < \infty$$

Thus, we can exchange the sum and the differentiation, and get

$$\begin{aligned} \frac{d^{n+1}}{dt^{n+1}} P_{xy}(t) &= \sum_{z=0}^{\infty} q^{(n)}(x, z) \frac{d}{dt} P_{zy}(t) \\ &= \sum_{z=0}^{\infty} q^{(n)}(x, z) \sum_{k=0}^{\infty} q(z, k) P_{ky}(t) \\ &= \sum_{k=0}^{\infty} P_{ky}(t) \sum_{z=0}^{\infty} q^{(n)}(x, z) q(z, k) \\ &= \sum_{k=0}^{\infty} q^{(n+1)}(x, k) P_{ky}(t). \end{aligned}$$

□

5. Minimal quasi-stationary distributions

So

$$\begin{aligned}
\frac{d^n}{dt^n} P_{x0}(t) \big|_{t=0} &= q^{(n)}(x, 0) \\
&= \sum_{z=0}^{\infty} q^{(n-1)}(x, z) q(z, 0) \\
&= q^{(n-1)}(x, 1) q(1, 0) \\
&= \sum_{z_1, \dots, z_{n-2} \in \mathbb{N}} q(x, z_1) q(z_1, z_2) \cdots q(z_{n-2}, 1) q(1, 0).
\end{aligned}$$

So

$$\frac{d^n}{dt^n} P_{n0}(t) \big|_{t=0} = q(n, n-1) q(n-1, n-2) \cdots q(1, 0),$$

and

$$\frac{d^n}{dt^n} P_{z0}(t) \big|_{t=0} = 0 \text{ if } z > n. \quad (5.20)$$

Now if v is a probability distribution on \mathbb{N} , with similar arguments used for proving (5.18) we can show that the n -th derivative of $F^v(t) = \sum_{z=1}^{\infty} v(z) P_z(X(t) = 0) = \sum_{z=1}^{\infty} v(z) P_{z0}(t)$ exists and is given by

$$\frac{d^n}{dt^n} F^v(t) = \sum_{z=1}^{\infty} v(z) \frac{d^n}{dt^n} P_{z0}(t). \quad (5.21)$$

In detail, this works by induction over n as follows:

The case $n = 0$ is clear. So suppose that (5.21) holds for some $n \in \mathbb{N}_0$. Then

$$\begin{aligned}
\frac{d^{n+1}}{dt^{n+1}} F^v(t) &= \frac{d}{dt} \left(\frac{d^n}{dt^n} F^v(t) \right) \\
&= \frac{d}{dt} \left(\sum_{z=1}^{\infty} v(z) \frac{d^n}{dt^n} P_{z0}(t) \right).
\end{aligned}$$

Now we would like to exchange the sum and the differentiation again, so we need to apply Theorem A.6. Thus we need to show that

$$\sum_{z=1}^{\infty} \left\| \frac{d}{dt} v(z) \frac{d^n}{dt^n} P_{z0}(t) \right\| < \infty.$$

$$\begin{aligned}
\sum_{z=1}^{\infty} \left\| \frac{d}{dt} v(z) \frac{d^n}{dt^n} P_{z0}(t) \right\| &= \sum_{z=1}^{\infty} v(z) \left\| \sum_{k=0}^{\infty} q^{(n+1)}(z, k) P_{k0}(t) \right\| \\
&\leq \sum_{z=1}^{\infty} v(z) \sum_{k=\max\{0, z-n-1\}}^{z+n+1} \|q^{(n+1)}(z, k)\| \\
&\leq \sum_{z=1}^{\infty} v(z) 2(n+1) 3^n C^{n+1} < \infty.
\end{aligned}$$

5. Minimal quasi-stationary distributions

So we can indeed exchange the sum and the differentiation and continuing the calculation we get

$$\frac{d^{n+1}}{dt^{n+1}} F^v(t) = \sum_{z=1}^{\infty} v(z) \frac{d^{n+1}}{dt^{n+1}} P_{z0}(t).$$

and that completes the induction step.

Let μ be another probability distribution on \mathbb{N} with $F^v = F^\mu$. We need to show that $\mu = v$. We note that (5.21) evaluated at $t = 0$. is just the finite sum

$$\frac{d^n}{dt^n} F^v(t) \big|_{t=0} = \sum_{z=1}^n v(z) \frac{d^n}{dt^n} P_{z0}(t) \big|_{t=0}. \quad (5.22)$$

Further for all $k \geq 1$ we have $\frac{d^k}{dt^k} P_{k0}(t) \big|_{t=0} = q(k, k-1) \cdots q(1, 0) > 0$. So we get

$$v(1) \frac{d}{dt} P_{10}(t) \big|_{t=0} = \frac{d}{dt} F^v(t) \big|_{t=0}.$$

And it follows from this that

$$v(1) = \frac{\frac{d}{dt} F^v(t) \big|_{t=0}}{\frac{d}{dt} P_{10}(t) \big|_{t=0}} = \frac{\frac{d}{dt} F^\mu(t) \big|_{t=0}}{\frac{d}{dt} P_{10}(t) \big|_{t=0}} = \mu(1).$$

Then if we suppose for induction hypothesis that for $n \in \mathbb{N}$ we have $\mu(k) = v(k)$, $k = 1, \dots, n$, then from

$$\frac{d^{n+1}}{dt^{n+1}} F^v(t) \big|_{t=0} = \sum_{z=1}^{n+1} v(z) \frac{d^{n+1}}{dt^{n+1}} P_{z0}(t) \big|_{t=0}$$

follows that

$$\begin{aligned} v(n+1) &= \frac{\frac{d^{n+1}}{dt^{n+1}} F^v(t) \big|_{t=0} - \sum_{z=1}^n v(z) \frac{d^{n+1}}{dt^{n+1}} P_{z0}(t) \big|_{t=0}}{\frac{d^{n+1}}{dt^{n+1}} P_{n+1,0}(t) \big|_{t=0}} \\ &= \frac{\frac{d^{n+1}}{dt^{n+1}} F^\mu(t) \big|_{t=0} - \sum_{z=1}^n \mu(z) \frac{d^{n+1}}{dt^{n+1}} P_{z0}(t) \big|_{t=0}}{\frac{d^{n+1}}{dt^{n+1}} P_{n+1,0}(t) \big|_{t=0}} \\ &= \mu(n+1). \end{aligned}$$

□

Corollary 5.11. *A probability distribution μ on \mathbb{N} is a quasi-stationary distribution if and only if the distribution of the absorption time F^μ is exponential. Furthermore for any $\theta > 0$ there exists at most one quasi-stationary distribution μ with $E_\mu R = \theta$. In particular, for $\theta = \frac{1}{\lambda_0} > 0$ there exists at most one quasi-stationary distribution μ with $E_\mu R = \frac{1}{\lambda_0}$ which means that there is at most one minimal quasi-stationary distribution.*

6. Acknowledgments

Proof. If μ is a quasi-stationary distribution, we know from the proof of the easy direction of Theorem 4.6 that F^μ is exponential.

On the other hand, if F^μ is exponential, then Lemma 5.9 yields $\mathcal{M}_\theta = \{\mu\}$ where $\theta = E_\mu R$. Theorem 4.27 finally gives that μ is a quasi-stationary distribution. \square

Proof of Theorem 5.8. Just as in the proof of Theorem 4.30 the family $\{\Phi^n \delta_x\}$ is tight. So after the Theorem of Prohorov, $\{\Phi^n \delta_x\}$ is weakly relative sequentially compact, and that means that every subsequence $(\Phi^{n_k} \delta_x)_{k \in \mathbb{N}}$ has a subsequence $(\Phi^{n_{k_j}} \delta_x)_{j \in \mathbb{N}}$ that converges to some probability measure μ . Because of Theorem 5.1 these limits are minimal quasi-stationary distributions and because of the previous corollary, the minimal quasi-stationary distribution is unique. With Lemma A.5 follows the assertion. \square

6. Acknowledgments

The author would like to thank Prof. V. Wachtel for proposing the interesting topic.

A. Appendix

Theorem A.1 (Monotone class theorem). *Let (Ω, \mathcal{F}) be a measurable space and let $P \subseteq \mathcal{F}$ be \cap -stable with $\Omega \in P$. Suppose \mathcal{H} is a vector space of random variables satisfying the following properties:*

(i) $A \in P$ implies $\mathbf{1}_A \in \mathcal{H}$.

(ii) $X_n \in \mathcal{H}, n \in \mathbb{N}$ and X bounded and $X_n \nearrow X$ implies $X \in \mathcal{H}$.

Then \mathcal{H} contains all $\sigma(P)$ -measurable random variables.

Proof. This follows from the fact, that since P is \cap -stable the σ -algebra generated by P is equal to the Dynkin-System generated by P and

$$\mathcal{D} = \{A \in \mathcal{F} | \mathbf{1}_A \in \mathcal{H}\}$$

is a Dynkin-System with $P \subseteq \mathcal{D}$. \square

Theorem A.2. *Let $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, g \neq 0$ satisfy the functional equation*

$$g(t+s) = g(t) \cdot g(s). \quad (\text{A.1})$$

and assume further that g is either continuous or monotone. Then $g(t) = e^{Ct}$ for some arbitrary constant $C \in \mathbb{R}$. If further $g \leq 1$, then C must be smaller or equal 0.

Proof. If there would be some $x_0 \geq 0$ with $g(x_0) = 0$, then for all $x \geq x_0$, $g(x) = g(x_0 + (x - x_0)) = g(x_0)g(x - x_0) = 0$. Since $0 = g(x_0) = g\left(\frac{x_0}{2}\right)^2$, also $g\left(\frac{x_0}{2}\right) = 0$ and by induction $g\left(\frac{x_0}{2^n}\right) = 0$ for all $n \in \mathbb{N}$. Since $\frac{x_0}{2^n} \xrightarrow{n \rightarrow \infty} 0$, $g(x) = 0$ for all $x \geq 0$, which contradicts $g \neq 0$.

A. Appendix

If there would be some $x_0 \geq 0$ with $g(x_0) < 0$, then $0 > g\left(\frac{x_0}{2}\right)^2$ which is a contradiction. So it must be $g > 0$. Thus we can define the map $f = \ln \circ g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$. Then f satisfies the functional equation $f(t+s) = f(t) + f(s)$. We show that f must satisfy $f(q) = q \cdot C$, $q \in \mathbb{Q}$ for some arbitrary constant $C \in \mathbb{R}$. $f(1) = f(1+0) = f(1) + f(0)$ shows that $f(0) = 0$. Every $q \in \mathbb{Q}$, $q > 0$ has the form $q = \frac{a}{b}$ where $a, b \in \mathbb{N}$. Then we have $qf(1) = \frac{a}{b}f\left(\frac{b}{b}\right) = af\left(\frac{1}{b}\right) = f\left(\frac{a}{b}\right) = f(q)$. $f(1) = C$ can be chosen arbitrarily. Now if g is either monotone or continuous, so is f either monotone or continuous.

1. Assume g continuous. Let $x \in \mathbb{R}_{\geq 0}$, then there exists a sequence $(q_n)_{n \in \mathbb{N}} \subseteq \mathbb{Q}_{\geq 0}$ with $\lim_{n \rightarrow \infty} q_n = x$, so $f(x) = \lim_{n \rightarrow \infty} f(q_n) = \lim_{n \rightarrow \infty} Cq_n = Cx$.
2. Assume g is monotone, without loss of generality monotone increasing. Let $x \in \mathbb{R}_{\geq 0}$, then there exists a sequences $(q_n)_{n \in \mathbb{N}} \subseteq \mathbb{Q}_{\geq 0}$ and $(r_n)_{n \in \mathbb{N}} \subseteq \mathbb{Q}_{\geq 0}$ with $q_n \searrow x$ and $r_n \nearrow x$ as $n \rightarrow \infty$. So $f(x) \geq \lim_{n \rightarrow \infty} f(r_n) = \lim_{n \rightarrow \infty} Cr_n = Cx$ and $f(x) \leq \lim_{n \rightarrow \infty} f(q_n) = \lim_{n \rightarrow \infty} Cq_n = Cx$, so in total $f(x) = Cx$ for all $x \geq 0$.

So $g(x) = e^{f(x)} = e^{Cx}$. If $g \leq 1$, then it's clear that $C \leq 0$. □

Theorem A.3. *Let (M, d) be a complete metric space and $a : \mathbb{N} \times \mathbb{N} \rightarrow M$ be a double-sequence of elements $a_{n,p} \in M$. Assume the following limits exist*

$$v(n) = \lim_{p \rightarrow \infty} a_{n,p}, \text{ for all } n \in \mathbb{N}, \quad (\text{A.2})$$

$$u(p) = \lim_{n \rightarrow \infty} a_{n,p}, \text{ for all } p \in \mathbb{N}, \quad (\text{A.3})$$

and assume further that either the limit (A.2) is reached uniformly in n or that the limit (A.3) is reached uniformly in p . Then the sequences $(v(n))_{n \in \mathbb{N}}$ and $(u(p))_{p \in \mathbb{N}}$ converge, and it holds

$$\lim_{n \rightarrow \infty} v(n) = \lim_{n \rightarrow \infty} \lim_{p \rightarrow \infty} a_{n,p} = \lim_{p \rightarrow \infty} \lim_{n \rightarrow \infty} a_{n,p} = \lim_{p \rightarrow \infty} u(p).$$

Proof. See any book about elementary analysis. □

Proposition A.4 (Kingman, 1963). *Let $f : (0, \infty) \rightarrow \mathbb{R}$ be continuous, and suppose that for each $h > 0$, the limit*

$$L(h) = \lim_{n \rightarrow \infty} f(nh)$$

exists and is finite. Then $L(h)$ is independent of h , and if we denote the common limit by L , then

$$\lim_{t \rightarrow \infty} f(t) = L.$$

Proof. See [2] Proposition 6.5.5. on page 225. □

Theorem A.5. *Let (M, d) be a metric space and $(a_n)_{n \in \mathbb{N}}$ be a sequence in M . Assume there exists $a \in M$ such that every subsequence $(a_{n_k})_{k \in \mathbb{N}}$ has a convergent sub-subsequence $(a_{n_{k_j}})_{j \in \mathbb{N}}$ with $\lim_{j \rightarrow \infty} a_{n_{k_j}} = a$. Then a_n converges to a .*

A. Appendix

Proof. Trivial. □

Theorem A.6. *Let $f_n : I \rightarrow \mathcal{C}$, $n \in \mathbb{N}$ be differentiable functions. If*

1. $\sum_{n=1}^{\infty} f_n$ converges pointwise on I .
2. $\sum_{n=1}^{\infty} f'_n$ converges absolutely on I .

Then the function $\sum_{n=1}^{\infty} f_n$ is differentiable on I and has derivate

$$f' = \sum_{n=1}^{\infty} f'_n.$$

References

- [1] N. I. Akhiezer. *The classical moment problem and some related questions in analysis*. Translated by N. Kemmer. Hafner Publishing Co., New York, 1965.
- [2] William J. Anderson. *Continuous-Time Markov Chains - An Applications-Oriented Approach*. Springer-Verlag, New York, 2011.
- [3] Søren Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [4] Heinz Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, fifth edition, 2002.
- [5] Leo Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.
- [6] Robert Cauty. Solution du problème de point fixe de Schauder. *Fund. Math.*, 170(3):231–246, 2001.
- [7] Kai Lai Chung. *A course in probability theory*. Academic Press, Inc., San Diego, CA, third edition, 2001.
- [8] Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-stationary distributions*. Probability and its Applications (New York). Springer, Heidelberg, 2013. Markov chains, diffusions and dynamical systems.
- [9] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [10] P. A. Ferrari, H. Kesten, S. Martinez, and P. Picco. Existence of quasi-stationary distributions. A renewal dynamical approach. *Ann. Probab.*, 23(2):501–521, 1995.
- [11] Pablo A. Ferrari, Servet Martínez, and Pierre Picco. Existence of nontrivial quasi-stationary distributions in the birth-death chain. *Adv. in Appl. Probab.*, 24(4):795–813, 1992.
- [12] W. L. Harkness and R. Shantaram. Convergence of a sequence of transformations of distribution functions. *Pacific J. Math.*, 31:403–415, 1969.
- [13] Harry Kesten. A ratio limit theorem for (sub) Markov chains on $\{1, 2, \dots\}$ with bounded jumps. *Adv. in Appl. Probab.*, 27(3):652–691, 1995.
- [14] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin Heidelberg New York, 2 edition, 2008.

References

- [15] M. G. Nair and P. K. Pollett. Correction: “On the relationship between μ -invariant measures and quasi-stationary distributions for continuous-time Markov chains”. *Adv. in Appl. Probab.*, 25(3):717–719, 1993.
- [16] M. G. Nair and P. K. Pollett. On the relationship between μ -invariant measures and quasi-stationary distributions for continuous-time Markov chains. *Adv. in Appl. Probab.*, 25(1):82–102, 1993.
- [17] J. Neveu. *Discrete-parameter martingales*. North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, revised edition, 1975. Translated from the French by T. P. Speed, North-Holland Mathematical Library, Vol. 10.
- [18] Anthony G. Pakes. Absorbing Markov and branching processes with instantaneous resurrection. *Stochastic Process. Appl.*, 48(1):85–106, 1993.
- [19] Erik A. van Doorn. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. in Appl. Probab.*, 23(4):683–700, 1991.
- [20] Erik A. van Doorn and Pauline Schrijner. Ratio limits and limiting conditional distributions for discrete-time birth-death processes. *J. Math. Anal. Appl.*, 190(1):263–284, 1995.

Erklärung zur Hausarbeit gemäß § 29 (Abs.6) LPO I

Hiermit erkläre ich, dass die vorliegende Hausarbeit von mir selbstständig verfasst wurde und dass keine anderen als die angegebenen Hilfsmittel benutzt wurden. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen sind, sind in jedem einzelnen Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht. Diese Erklärung erstreckt sich auch auf etwa in der Arbeit enthaltene Zeichnungen, Kartenskizzen und bildliche Darstellungen.

Ort, den Datum